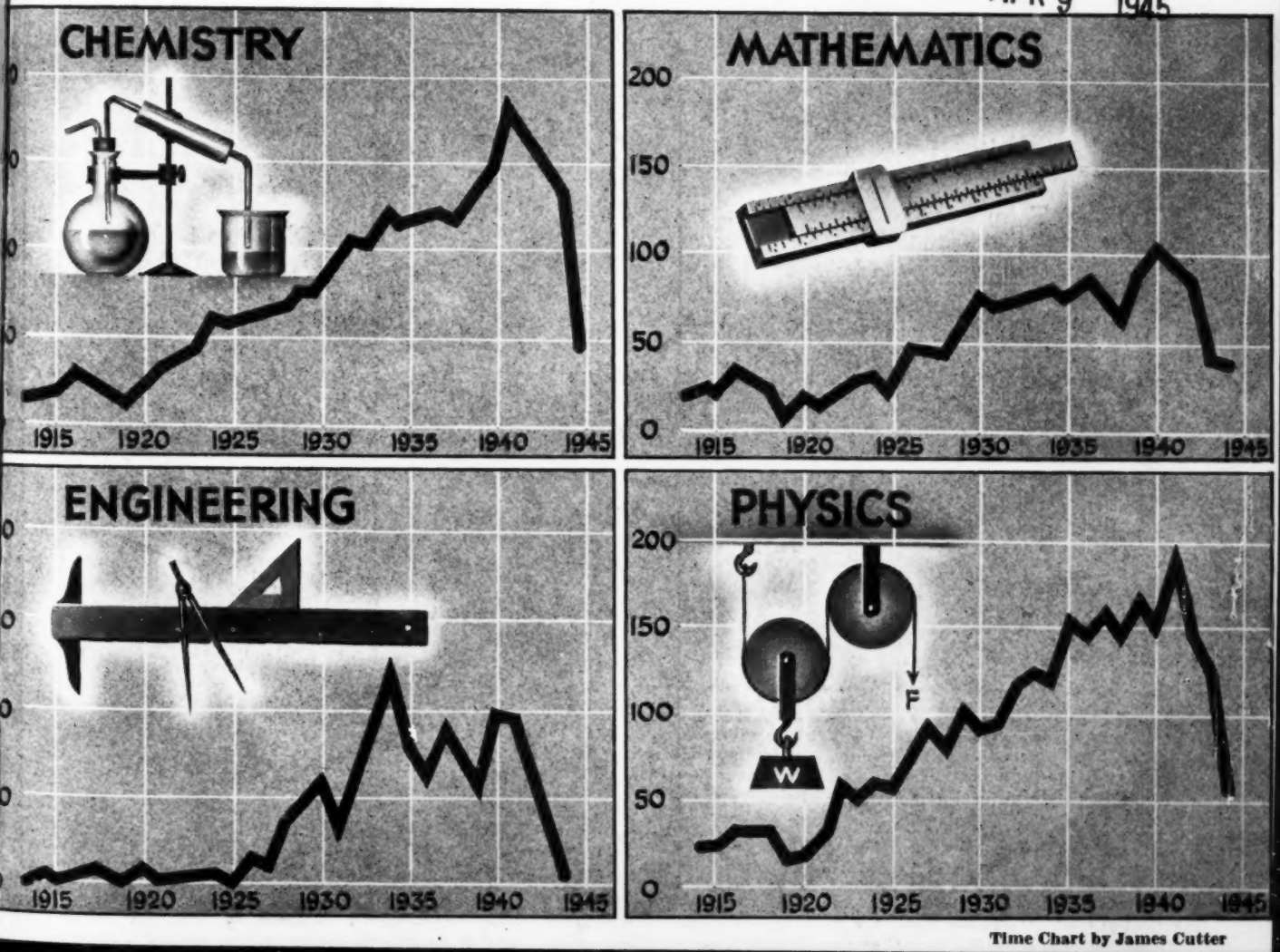
JOURNAL
OF
APPLIED PHYSICS

ics Lib.

"The Road to Scientific Bankruptcy"

APR 9 1945



Time Chart by James Cutter

JOURNAL
OF
APPLIED PHYSICS

Vol. 16, No. 3 MARCH, 1945

In This Issue

Page

Electrical Research in the U. S. S. R. GERALD OSTER 121

Contributed Original Research

General Theory of Stresses and Displacements in Layered Soil Systems. II D. M. BURMISTER 126

Electrical Charging of Electron Diffraction Specimens D. G. BRUBAKER AND M. L. FULLER 128

100-kv Electron Microscope L. MARTON 131

Solution of the Diffusion Equation Applicable to the Edgewise Growth of Pearlite W. H. BRANDT 139

Production Histories of Oil Producing Gas-Drive Reservoirs MORRIS MUSKAT 147

Effects of Temperature and Humidity on the Physical Properties of Tire Cords J. H. DILLON AND I. B. PRETTYMAN 159

Numerical Solution of Ordinary and Partial Differential Equations by Means of Equivalent Circuits GABRIEL KRON 172

Erratum: Physical Aspects of Foaming in Steam Generation HANS M. CASSEL 186

Letter to the Editor W. JAMES LYONS 187

In Every Issue

Here and There 125

New Book 125

Recent Applications of Physics 188

Index to Advertisers xxviii

BOARD OF EDITORS

ELMER HUTCHISSON, *Editor*

Associate Editors

R. B. BARNES
OTTO BEECK
W. F. BUSSE
S. H. CALDWELL
CLARK GOODMAN
SELIG HECHT
W. J. LYONS
MORRIS MUSKAT
THOMAS H. OSGOOD
PAUL E. SABINE
FREDERICK SEITZ
J. R. TOWNSEND

Publication of the American Institute of Physics

ADMINISTRATIVE STAFF

HENRY A. BARTON, *Director*

MADELINE M. MITCHELL, *Publications Manager*

JOHN T. TATE, *Adviser on Publications*

GOVERNING BOARD

PAUL E. KLOPSTEG, *Chairman*

GEORGE B. PEGRAM, *Secretary*

T. D. COPE
R. B. DOW
L. A. DUBRIDGE
FLOYD A. FIRESTONE
HARVEY FLETCHER
R. C. GIBBS
K. S. GIBSON
HERMANN MARK
A. G. WORTHING
JOHN T. TATE
L. W. TAYLOR
N. W. TAYLOR
WALLACE WATERFALL

The JOURNAL OF APPLIED PHYSICS, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as PHYSICS. The JOURNAL OF APPLIED PHYSICS publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case School of Applied Science, Cleveland 6, Ohio.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price

	U. S. AND CANADA	ELSEWHERE
To members of scientific societies	$5.00	$5.70
To all others	7.00	7.70

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 15, 1944—**$94.50**
Yearly back number rate: **$7.70**
Single copies: **$0.70** each.

Changes of address and proofs should be addressed to the Publications Manager.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879
Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1945, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 16, Number 3 March, 1945

Electrical Research in the U.S.S.R.

BY GERALD OSTER

*Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts**

(Received September 16, 1944)

INTRODUCTION

THE brilliant successes of the Red Army in the field of battle did not come as a surprise to those people who were acquainted with the growth of Soviet industry. This industry is backed, as in the United States, by a great amount of scientific research.

Soviet scientists attack their problems in much the same manner as do scientists the world over. Although their organizational set-up is somewhat different from that of this country, many of their scientific institutes are not unlike some of our own research laboratories. For example, the Bell Telephone Research Laboratory with its large staff of engineers, physicists, chemists, and mathematicians, all working on one broad problem, is very much like one of the larger Soviet research institutes.[1]

Too few Americans are familiar with the important contributions to electrical research made by Soviet scientists. This paper is an attempt to acquaint American scientists with the more outstanding work in Soviet electrical research. It is important for us to become familiar with this work, not only to supplement our own research, but also to inform us as to the potential market the Soviet Union can become for our electrical products. The U.S.S.R. is destined to become one of our greatest foreign markets for electrical goods.[2]

* Present address.

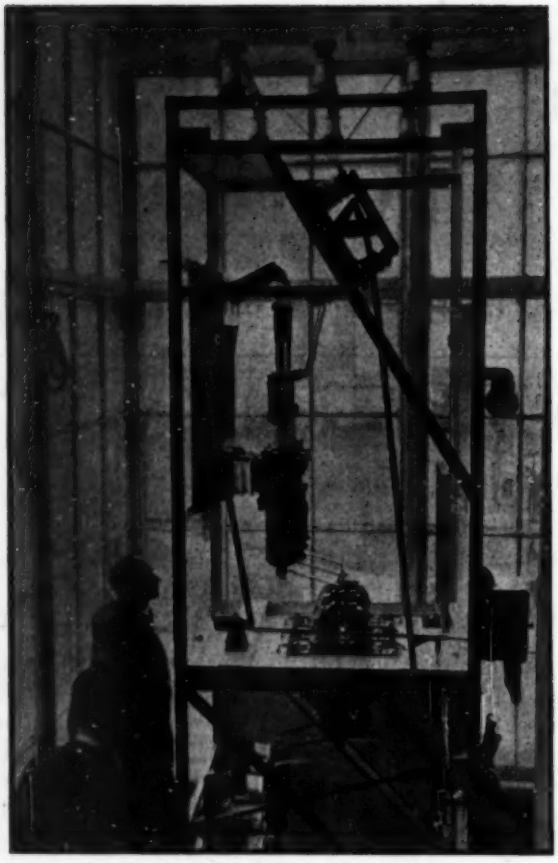
SEMICONDUCTOR AND INSULATION RESEARCH

Mainly because of the efforts of A. Joffe and his school, the Soviet Union has achieved a high place in the field of semiconductor research. The claimed ratings[3] for Soviet rectifiers and for their photoelectric cells place them in the same class as the United States and Germany in this field. Of particular interest is their stable infra-red sensitive photo-voltaic cell with a sensitivity of 10,000 microamperes per lumen. This is a sufficiently high rating to warrant their possible use as converters of solar energy into electrical energy on an industrial scale.

Much of their work on semiconductors is accompanied by theoretical work done by such competent physicists as J. Frenkel and L. Landau, both well known in this and other branches of physics.

A. Joffe, together with A. F. Walther, has made fundamental experimental and theoretical studies on the nature of electrical breakdown.[4] Joffe, Walther, Inge, Wul, and Goldman have performed interesting experiments on breakdown in crystals[5] and gases.[6] This work has, as in the case of research on semiconductors, been accompanied by theoretical interpretations.[7]

Out of this work have come new industrial insulators such as aluminum oxide glazes; also, new methods of impregnation of paper and of glass cloth with cable oils under high pressures have been developed. The efforts of these scientists will be applied to the projected dam at Kuibishev, on the Volga River, where a hydro-



Sovfoto

FIG. 1. A 560-kilowatt transformer turned out by the Ultraelectromashina Works (Sverdlovsk), as exhibited in the Commissariat of the Building Industry of the U.S.S.R., Moscow.

electric station will be constructed having a capacity "twice the power of the TVA, Grand Coulee, and Bonneville developments combined."[8] The power from this dam will be carried by transmission lines at 440,000 volts to Moscow, 500 miles away, and to the Urals, over 300 miles away.

Accompanying Joffe's work on the dielectric breakdown of materials is his construction of unique types of high voltage generators. In a recent article[9] Joffe and Hochberg discuss some of the limitations of the van de Graaff generator. They present designs of their multidisk generator, which also generates several million volts but is calculated to give ten thousand times more power per unit volume. This suggests its use in industry and also suggests the possibility of constructing reasonably sized, very high voltage motors.

Investigations, over a period of many years, on the amorphous state (plastics, liquids, glasses, etc.) have been carried out by P. Kobeko and his school. His studies on the dielectric constant, dielectric loss, and mechanical properties of insulating materials are described in articles appearing in the journal *Technical Physics of the U.S.S.R.*

Kobeko has collaborated with I. Kurtchatov[10] in examining the phenomena of the ferroelectric characteristics of crystals of the Rochelle salt type. Kurtchatov[11] has extended this work and has presented a theory of ferroelectricity.

NON-LINEAR OSCILLATIONS RESEARCH

One of the most original contributions of Soviet scientists to electrical research is their work in the field of non-linear oscillations. I. Mandelstam and N. Papalexi are the leaders in this field. Mandelstam, who has made important contributions in various fields of physics, is best known for his discovery (with Landsberg) of the Raman effect in crystals just before the effect was discovered by Raman in 1928. Before 1934 Mandelstam and Papalexi had succeeded in constructing alternating current generators by periodically altering a parameter, such as the capacitance of a condenser, in an inductance-capacitance circuit, the inductance being a non-linear element such as a saturated core. The generators described[12] are based on this principle. They have important advantages over the conventional type dynamos and are much lighter than an ordinary generator of similar power. They use much less metal and yet have the same efficiency. These generators may be of use as voltage sources in mobile x-ray units.

The Heaviside calculus, which is applied (perhaps indiscriminately, as modern electrical engineers now realize) to ordinary "linear" circuits, is not at all applicable to a circuit with variable parameters. Then the equations are non-linear and are much more difficult to solve. Some of the most accomplished Soviet mathematicians, working in close collaboration with their engineers, have made many contributions in this respect.[13]

Soviet scientists are further pursuing this field and have applied the theory of non-linear vibrations to many new vacuum-tube circuits. It does not seem unreasonable that the theory of non-linear oscillations will be a working tool of radio scientists in the future.

ELECTRONICS AND RADIO RESEARCH

Physicists in the U.S.S.R. have made detailed theoretical and experimental studies on the

nature of the photoelectric effect and thermionic emission. The surface studies of I. Tamm and P. Lukirsky[14] are outstanding.

There is also much work being done in the U.S.S.R. on electrical discharge in gases with applications to high voltage vacuum tubes (as high as one million volts) and to switch gear and other industrial equipment.[15] These efforts have led to the development of large ionic current sources, notably the so-called gasomagnetron which produces much more ionic current than is available from canal rays.[16]

The construction of radio tubes, magnetrons, etc., in the Soviet Union has usually been accompanied by theoretical work.[17] Similarly, radio transmission theory, notably that applied to anisotropic media (i.e., land, sea, and air as encountered by ship to shore signalling), has been studied.[18]

V. Arkadiev has performed some interesting experiments in the photography of radio waves. Copper and zinc pellets are placed under a damp paper containing phenolphthalein. By the voltaic effect of the dissimilar metals, a trace of the radio wave (of any wave-length) appears on the paper.[19] Arkadiev has recently announced[20] that his laboratory has obtained screens which fluoresce when impinged directly with radio waves. This work is a result, in part, of the efforts of Soviet physicists, especially S. Vavilov,[21] in advancing our knowledge of the nature of fluorescence and luminescence.

The gap in the spectrum between microwaves and heat waves was bridged some years ago by Glogoleva-Arkadiev. She was able to produce electromagnetic waves from the micro region down to hundreths of a millimeter by her "mass radiator."[22] This method consists of interposing in the spark gap of the exciting induction coil a paste of fine metal particles mixed with heavy mineral oil. The wave-length is determined by the dimensions of the particles and the radiation is monochromatic if the particles are of uniform size. The optical properties of this radiation were studied with wire mesh gratings.

In still another branch of electricity Mandelstam and Papalexi have made important contributions: namely, in the field of radio wave interference.[23] The method consists in applying some of the output of a fixed frequency transmitter across two plates of an oscilloscope. This signal is received some distance away, is amplified, multiplied in frequency by an integral value of the incoming signal, and is sent back to the original transmitter and impressed across the other plates of the oscilloscope. By varying continuously the frequency of the first transmitter and noting the change in phase by observing the change in the Lissajous figure produced on the oscilloscope, the velocity of propagation of the radio wave and the distance between the two transmitters can be determined. This technique was employed to determine the velocity of radio waves over water and various soils at different distances from the transmitter. It has been used since 1931 as a distance meter and is employed extensively by the Soviets in making geodesic surveys of their Arctic territory.[24] Various modifications of the method have been made to study diffraction and dispersion phenomena, including its use in meteorological observations and its use in the determination of the velocity of moving objects.

Sovfoto

FIG. 2. A Russian radio technician in the Moscow Sound Registering Factory operating an American-made (RCA) amplifier.

RESEARCH IN MAGNETISM

Many American scientists are already familiar with the work of L. Landau on the theory of diamagnetism, N. Arkulov on the theory of the hystereses of ferromagnets, and of the achievement by P. Kapitza of very high magnetic field strengths (above 300,000 gauss). Also, the work of Soviet scientists on low temperature studies of magnetism and their research on superconductivity is well known. In recent years engineers and physicists in the U.S.S.R. have applied the theory of magnetism to defectoscopy, the detection of flaws in metal objects. V. K. Arkadiev[25] has helped develop instruments which quickly detect flaws in ferromagnetic manufactured articles. Supersonic defectoscopy developed by S. Sokolov[26] is widely used in Soviet factories to examine non-metallic as well as metallic objects.

Work is now going on under the direction of V. Arkadiev and N. Arkulov on magnetic dispersion at microwave frequencies to study the magnetic analogy for ferromagnets of the Debye dispersion in dielectrics.

CONCLUSIONS

This survey of some of the major accomplishments of the Soviet Union in electrical research may serve to indicate the directions in which they have been working. It is apparent that their research achieves the same high standard of the electrical research carried on in the United States.

The future holds many exciting possibilities of collaboration with Soviet scientists. In the words of Professor E. O. Lawrence: "All American scientists have a high regard for their Russian colleagues and have long felt that closer contacts would accelerate scientific progress to the benefit of all. During the war communications are necessarily restricted, but let us do everything we can to bring about, after the war, the fullest cooperation and even more cordial relations."[27]

BIBLIOGRAPHY

(1) For an excellent account of Soviet laboratories see J. G. Crowther, *Soviet Science* (E. P. Dutton & Co., Inc., New York, 1936).

(2) Our postwar trade with the U.S.S.R. has been estimated to amount to ten billion dollars. The Soviet Union has already bought from us, via lend lease, a million miles of wire and one-quarter million telephone sets.

(3) For a general survey of their achievements in this field, see A. F. Joffe, Vestnik Akad. Nauk S.S.S.R. (in Russian) **10**, 12 (1940); and S. J. Lukyanov, Zhurn. Tech. Fizik. (in Russian) **8**, 3 (1943).

(4) A. Joffe, *The Physics of Crystals* (McGraw-Hill Book Company, Inc., New York, 1928); A. Walther and L. Inge, Tech. Phys. U.S.S.R. **5**, 335 (1938).

(5) E.g., Joffe, Walther, Inge, Wul, and Goldman, Tech. Phys. U.S.S.R. **3**, 16 (1936) and **5**, 561 (1938).

(6) Meeting of the bureau of electrical insulation, Tech. Phys. U.S.S.R. **4**,752 (1937).

(7) J. Frenkel, Tech. Phys. U.S.S.R. **5**, 685 (1938).

(8) Wendell Willkie, *One World* (Simon and Schuster, New York, 1943), p. 29. The exact figures on completion are: an installed capacity of 4,000,000 kw with an annual output of over 7,000,000,000 kw-hr. See A. V. Vinter, *Twenty-Five Years of Power Development in the U.S.S.R.* (The American Russian Institute, New York, 1943).

(9) A. Joffe and B. Hochberg, J. Phys. U.S.S.R. **2**, 243 (1940).

(10) P. Kobeko and I. Kurtchatov, Zeits. f. Physik **66**, 192 (1930).

(11) I. V. Kurtchatov, *Le Champ Moleculair Dans Les Dielectriques* (Hermann et Cie., Paris, 1936).

(12) L. Mandelstam, N. Papalexi, *et al.*, Tech. Phys. U.S.S.R. **2**, 81 (1935).

(13) N. Krylov and N. N. Bogoliuboff, *New Methods in Non-Linear Mechanics—Their Application to the Study of Electronic Generators* (in Russian) (Moscow, 1934); A. A. Andranov and S. E. Khaikin, *Theory of Vibrations* (in Russian) (Leningrad, 1937); a translation into English by S. Lefshetz of some work of N. Krylov and N. Bogoliuboff has recently appeared: *Introduction to Non-Linear Mechanics* (Princeton University Press, 1943).

(14) P. Lukirsky, Tech. Phys. U.S.S.R. **3**, 685 (1936).

(15) Conference on electrical discharge and its application, Tech. Phys. U.S.S.R. **5**, 561 (1938).

(16) I. Vigdorchick, J. Phys. U.S.S.R. **1**, 151 (1939).

(17) E.g., G. Grunberg, Tech. Phys. U.S.S.R. **2**, 482 (1935); **5**, 3 (1938).

(18) E.g., V. L. Ginsberg, J. Phys. U.S.S.R. **3**, 95 (1940).

(19) V. Arkadiev, Comptes rendus Acad. Sci. U.S.S.R. **3**, 412 (1934). See also J. Exp. Theor. Phys. U.S.S.R. **7**, 87 (1937).

(20) V. Arkadiev, Nature **154**, 157 (1944).

(21) E.g., S. Vavilov, Physik. Zeits. Sowjetunion **5**, 379 (1934).

(22) A. Z. Glogoleva-Arkadiev, Nature **113**, 640 (1924); M. Levitsky, Physik. Zeits. **35**, 361 (1934).

(23) I. Mandelstam and N. Papalexi, Tech. Phys. U.S.S.R. **4**, 767 (1937).

(24) J. Phys. U.S.S.R. **1**, 388, 393 (1939); **4**, 13 (1941).

(25) V. K. Arkadiev, Conference on magnetic defectoscopy, Tech. Phys. U.S.S.R. **4**, 754 (1937).

(26) S. Sokolov, Tech. Phys. U.S.S.R. **2**, 522 (1935).

(27) E. O. Lawrence, *Science in Soviet Russia* (Jaques Cattell Press, Lancaster, Pa. 1944), p. 89. This book is a collection of speeches by prominent American scientists. Also, in the same book is an article by V. K. Zworykin, "Soviet developments in electronics."

Here and There

Browder J. Thompson, associate research director of RCA Laboratories at Princeton, New Jersey, was killed in action during a flight in an Army airplane in the Mediterranean Theater of Operations July 4, 1944, while engaged in a special mission for the Secretary of War. Since December, 1943 he had been on leave of absence from RCA as Expert Consultant in the Office of the Secretary of War.

Recent new grants-in-aid announced by Dr. Robert C. Hockett, Scientific Director of The Sugar Research Foundation, bring the research budget of the Foundation to a total of more than a quarter of a million dollars. The Foundation maintains a laboratory at the Massachusetts Institute of Technology which constitutes a training center for carbohydrate chemists. Pure and applied research in the field of organic chemistry as it relates to the sugars is in progress. The Foundation's program of developing new uses for sugar in the world of tomorrow and in establishing the role of sugar in the human diet is being accomplished by grants-in-aid to universities and colleges throughout the United States.

Polytechnic Institute of Brooklyn announced in January the organization of a separate division of polymer chemistry under the direction of Dr. Herman F. Mark, who has been professor of organic chemistry for the past five years. With this move, Polytechnic becomes the first educational institution in the United States to set up a complete division for polymer chemistry.

The following have been elected officers of the Executive Committee of the Division of High-Polymer Physics of the American Physical Society for 1945: Chairman, W. F. Busse; Vice Chairman, S. D. Gehman; Secretary, W. J. Lyons; Treasurer, L. A. Wood. Other members of the Executive Committee are H. Mark, newly elected to the Committee, R. B. Barnes, a previous incumbent, and J. H. Dillon, the retired Chairman.

New Appointments

Clinton R. Hanna, manager of the electro-mechanical department of the Westinghouse Research Laboratories, has been appointed an Associate Director of the Laboratories, it was announced by Dr. L. W. Chubb, Director of Research. For his work in the development of the tank gun stabilizer, which enables Allied tanks to fire accurately while in motion, he was awarded a Presidential Citation in 1942. A holder of more than 80 patents, Mr. Hanna designed the silverstat, an automatic voltage regulator, and developed the photophone, one of the first successful methods of producing talking movies.

Born on December 17, 1899, in Indianapolis, Indiana, Mr. Hanna attended Purdue University where he was graduated in 1922 with the degree of bachelor of science in electrical engineering, receiving the professional degree of electrical engineer four years later. He is a member of the Institute of Radio Engineers and a fellow of the Acoustical Society of America and the American Institute of Electrical Engineers, and has contributed more than a score of papers to technical journals.

W. Myron Owen, who recently resigned the vice presidency of the Detroit Harvester Company, of which he remains a director, is the new president and co-owner of Aerovox Corporation, New Bedford, Massachusetts.

Science reports that Dr. Zay Jeffries, technical director of the lamp department at Cleveland of the General Electric Company, has been appointed, effective January 1, vice president of the company and manager of the chemical department.

Armour Research Foundation of Illinois Institute of Technology recently announced the election of Dr. J. E. Hobson as director. Dr. Hobson, previously director of the Department of Electrical Engineering at Illinois Institute of Technology, succeeded Mr. Harold Vagtborg, who resigned to become president of the Midwest Research Institute.

Vacuum Engineering Division of National Research Corporation has announced the appointment to its sales staff of E. C. Bowen, formerly Eastern Divisional Sales Manager for Central Scientific Company of Chicago. Mr. Bowen will devote his attention to industrial applications of low pressure.

New Book

Basic Radio

By C. L. Boltz. Pp. 272, Figs. 166. The Ronald Press Company, New York. Price $2.25.

This small book is an excellent elementary introduction to a vast subject. It starts with simple ideas—a battery and a wire—and advances rapidly to more complex ones. The author's style calls for the cooperation of his reader: he expects the reader to be an intelligent person who is willing to advance his knowledge by study. There is no coddling, but the speed of advance is well measured. There is constant appeal to observation and experiment and to simple computations; but there is also a refusal to become involved in mathematical difficulties or to be drawn aside into interesting historical byways. Seldom does an author set his aims so carefully and stay so religiously within the bounds which he has prescribed for himself.

This text was prepared for use with British Air Training Cadets and is one of "Nelson's Aeroscience Manuals." It is a splendid example of what can be accomplished in a few pages when there are specific objectives in view. In this case, the objective is to bridge the gap between no understanding of electricity, or at most a flimsy understanding, and a clear conception of the basic elements of radio operation. The reviewer can agree with the publisher's note that Mr. Boltz's book is "exceptional for clearness, simplicity, rapid rate of progress into the subject, as well as original and distinctive in the skillful building of a real mastery of each particular topic." One must admire the author's clear treatment of ideas, both simple and difficult. He shows great restraint at times, and his ability to ignore the irrelevant is remarkable.

The book is written for class or self-study. There are examples of application of principles frequently, and each chapter contains well-graded numerical problems for which answers are provided in the back of the book. Emphasis is upon practical knowledge and its application. The treatment of condensers (Chap. VI) seems to the reviewer as not up to the rest of the book in clarity or precision. The author uses enough of the facile but sometimes inaccurate language of the radio "ham" to suit the needs of certain readers and to serve as a bridge over the gap to more scientific terminology. The book is definitely on the scientific side, even though simple, and is not just another "ham" book.

Richard M. Sutton
Haverford College

Contributed Original Research

The General Theory of Stresses and Displacements in Layered Soil Systems. II

D. M. BURMISTER*
Department of Civil Engineering, Columbia University, New York, New York
(Received August 24, 1944)

The equations of stresses and displacements in a two-layer system are derived in this paper for Case 2 with a frictionless interface between the two layers.

INTRODUCTION

UNDER certain conditions with repeated heavy concentrations of airplane wheel loadings, experience seems to indicate that the continuity of shearing stress between the pavement layer and the subgrade of an airport runway may be partially or almost completely broken down and lost. The general equations for stresses and displacements are derived for a frictionless interface for the analysis of such problems. The assumptions and boundary conditions for layered systems were given in the previous paper on layered soil systems. Starting with Eqs. (6) and (7) of the former paper for the stresses and displacements in the two-layer system, the equations are derived for Case 2 with a frictionless interface.

1. BOUNDARY AND CONTINUITY CONDITIONS, CASE 2. FRICTIONLESS INTERFACE

(a) Boundary Conditions at the Surface of the Ground, $z=-h$

Distribution of surface loading,

$$\sigma_z=-mJ_0(mr). \tag{19a}$$

$$-mJ_0(mr)[A_1m^2e^{-mh}+B_1m^2e^{mh}-C_1m(1-2\mu_1+mh)e^{-mh}+D_1m(1-2\mu_1-mh)e^{mh}]=-mJ_0(mr).$$

Shearing stress at surface, $\tau_{rz}=0$. (19b)

$$mJ_1(mr)[A_1m^2e^{-mh}-B_1m^2e^{mh}+C_1m(2\mu_1-mh)e^{-mh}+D_1m(2\mu_1+mh)e^{mh}]=0.$$

(b) Continuity Conditions at the Interface

Normal stress, $\sigma_{z1}=\sigma_{z2}$. (20a)

$$[A_1m^2+B_1m^2-C_1m(1-2\mu_1)+D_1m(1-2u_1)]=[B_2m^2+D_2m(1-2\mu_2)].$$

Settlement $w_1=w_2$. (20b)

$$\frac{1+\mu_1}{E_1}[A_1m^2-B_1m^2-C_1m2(1-2\mu_1)-D_1m2(1-2\mu_1)]=\frac{1+\mu_2}{E_2}[-B_2m^2-D_2m2(1-2\mu_2)].$$

Shearing stress (20c)

$$\tau_{rz1}=0=[A_1m^2-B_1m^2+C_1m2\mu_1+D_1m2\mu_1],$$

$$\tau_{rz2}=0=[-B_2m^2+D_2m2\mu_2].$$

(c) Discontinuities at the Interface

Radial stress $\sigma_{r1}\neq\sigma_{r2}$.
Horizontal displacement $u_1\neq u_2$. (21)

2. TWO-LAYER SYSTEM. EQUATIONS OF STRESS AND DISPLACEMENT

The coefficients A_1, B_1, C_1, and D_1 were evaluated to satisfy the boundary and continuity conditions for a frictionless interface, as follows:

$$A_1m=[C_1(F-2\mu_1)-D_1(1-F)], \tag{22a}$$

$$B_1m=[C_1F-D_1(1-2\mu_1-F)], \tag{22b}$$

$$C_1m=[(1-2\mu_1-F)e^{mh}+(2\mu_1+mh)e^{mh}-(1-F)e^{-mh}]\frac{1}{\Delta}, \tag{22c}$$

* Publication assisted by the Ernest Kempton Adams Fund for Physical Research of Columbia University.

$$D_1 m = [Fe^{mh} - (F-2\mu_1)e^{-mh} - (2\mu_1 - mh)e^{-mh}]\frac{1}{\Delta}, \quad (22d)$$

where the common denominator

$$\Delta = [Fe^{2mh} + (2F-1)2mh - (1+2m^2h^2) + (1-F)e^{-2mh}], \quad (22e)$$

and where the coefficients of the strength properties of the two layers are:

$$n = \frac{E_2}{E_1}\left[\frac{1+\mu_1}{1+\mu_2}\right], \quad F = \left[\frac{(1-\mu_2)+n(1-\mu_1)}{2(1-\mu_2)}\right]. \quad (22f)$$

Substituting these values of the coefficients A_1, B_1, C_1, and D_1 into Eqs. (6) (previous paper) for layer No. 1, and simplifying, the general stress and displacement equations were obtained for the two-layer system with a frictionless interface.

(a) At the Surface of Layer No. 1

Settlement (23)

$$w = J_0(mr)\frac{2(1-\mu_1^2)}{E_1}\left[\frac{F-[2F-1-2mh]e^{-2mh}-(1-F)e^{-4mh}}{F+[(2F-1)2mh-(1+2m^2h^2)]e^{-2mh}+(1-F)e^{-4mh}}\right].$$

Approximate radius of curvature, $\frac{1}{R} = \frac{\partial^2 w}{\partial r^2}$, (24)

$$\frac{1}{R} = \left[m^2 J_0(mr) - \frac{m^2 J_1(mr)}{mr}\right]\left[\frac{(1-\mu_1^2)}{E_1}\right][\text{Bracket of Settlement Eq. (23)}].$$

(b) At the Interface Between Layers No. 1 and No. 2

Normal stress (25)

$$\sigma_z = -mJ_0(mr)[2F-1]\left[\frac{(1+mh)e^{-mh}-(1-mh)e^{-3mh}}{F+[(2F-1)2mh-(1+2m^2h^2)]e^{-2mh}+(1-F)e^{-4mh}}\right].$$

Radial stress (26)

$$\sigma_{r1} = mJ_0(mr)\left[\frac{[(1+mh)-2F(1-mh)]e^{-mh}-[(1+3mh)-2F(1+mh)]e^{-3mh}}{F+[(2F-1)2mh-(1+2m^2h^2)]e^{-2mh}+(1-F)e^{-4mh}}\right]$$

$$-\frac{mJ_1(mr)}{mr}\left[\frac{[(1-2\mu_1)(1+mh)-2F(1-2\mu_1-mh)]e^{-mh}+ -[(1-2\mu_1)(1+mh)+2mh-2F(1-2\mu_1+mh)]e^{-3mh}}{F+[(2F-1)2mh-(1+2m^2h^2)]e^{-2mh}+(1-F)e^{-4mh}}\right].$$

Settlement (27)

$$w = J_0(mr)\frac{2(1-\mu_1^2)}{E_1}\left[\frac{(1+mh)e^{-mh}-(1-mh)e^{-3mh}}{F+[(2F-1)2mh-(1+2m^2h^2)]e^{-2mh}+(1-F)e^{-4mh}}\right].$$

Approximate radius of curvature (28)

$$\frac{1}{R} = \left[m^2 J_0(mr) - \frac{m^2 J_1(mr)}{mr}\right]\left[\frac{2(1-\mu_1^2)}{E_1}\right][\text{Bracket of Settlement Eq. (27)}].$$

Equations (25) and (27) may be shown to be identical with those given by Marguerre at the interface.[1] If the modulus becomes infinite, that is, a frictionless rock surface at the base of layer No. 1, and Poisson's ratio, $\mu_1 = \mu_2 = \frac{1}{2}$ Eq. (25) for the normal stress at the rock surface becomes identical with that given by Biot.[2]

[1] K. Marguerre, Ing. Arch. **4**, 332 (1933).
[2] M. A. Biot, Physics **6**, 367 (1935).

The Electrical Charging of Electron Diffraction Specimens

D. G. Brubaker and M. L. Fuller
Research Division, Technical Department, The New Jersey Zinc Company, Palmerton, Pennsylvania
(Received September 30, 1944)

INTRODUCTION

In the use of electron diffraction for the identification of materials by the reflection method, difficulty is often encountered because the material acquires an electric charge under the influence of the electron beam which strikes it. A number of workers in this field have adopted the practice of reducing negative charges on specimens of such material by use of an auxiliary beam of low velocity electrons which is caused to strike the specimen surface. This practice was first brought to the writers' attention by J. E. Ruedy of the RCA Laboratories. The use of such an auxiliary beam on glass specimens was described by H. Kamogawa.[1] His explanation of its action was that the negative charge accumulated on the glass by absorption of high velocity electrons was balanced by emission of secondary electrons produced by the auxiliary beam of low velocity electrons. Although no other observations regarding the accumulated charge and the mechanism of its reduction in such cases have been published, so far as the writers are aware, conversations with other workers in electron diffraction have indicated the prevailing opinion that the accumulated charge is negative and that the mechanism of reduction as outlined by Kamogawa is essentially correct.

In the use of electron diffraction in this laboratory a wide variety of poorly conducting materials has been examined by the reflection method. Practically all of them acquired an electric charge during exposure to the 40-kv main electron beam, but in only two of the cases is the charge thought to have been negative. In all the many others the evidence indicates that the charge was positive. During the last two years an auxiliary low voltage electron beam has been used to irradiate specimens which tend to become charged, and it has been found possible to reduce the charge, whether positive or negative, to a negligible amount in all cases.

OBSERVATIONS ON CHARGING OF SPECIMENS

The geometrical relationships of the main electron beam, the specimen, and the diffracted beams in the making of diffraction patterns by the reflection method is indicated schematically in Fig. 1, which is not drawn to scale. The main

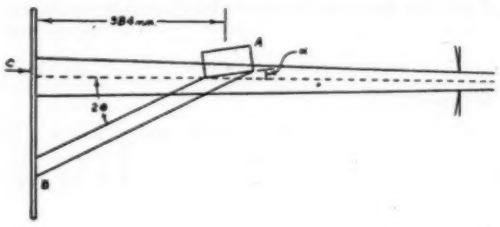


Fig. 1.

beam strikes the surface of specimen A at a grazing angle of the order of 0.5°. Diffracted electrons come off at angles 2θ and strike the photographic plate, as at B. After the plate has been exposed, we customarily withdraw the specimen from the beam and move the photographic plate a short distance horizontally before turning off the electron beam. The camera construction which makes this possible has been described.[2] This procedure causes the beam to trace a base line about 0.2 mm wide upon the plate. This is intended to mark the center of the diffraction arcs, and does so if the specimen was not electrically charged. Base lines are shown in the diffraction patterns illustrated in Fig. 2.

If there is an electric charge on a specimen it is indicated in some or all of three ways: (1) The trace on the plate of that part of the electron beam which passes the specimen is distorted from its normal shape and size, (2) the lines of the diffraction pattern are broadened, (3) the lines are shifted from their normal positions relative to that of the undeflected electron beam. In Fig. 2(a) is shown a pattern which illustrates all of these effects. It is the phenomenon of pattern shift which is most troublesome, since it makes accurate determination of the crystal interplanar spacings and subsequent identification nearly impossible, except in special cases.

This shifting of diffraction patterns was first recognized by us in 1940 during attempts to secure good diffraction patterns by the reflection method from relatively thick layers of such

[1] H. Kamogawa, Phys. Rev. **58**, 660 (1940).

[2] Metals and Alloys **10**, 85–89; **10**, 122–125 (1939).

substances as pigment ZnS and TiO_2. These became charged by the 40-kv electron beam. Measurements of the arc radii in the diffraction patterns, made with the recorded base line assumed passing through the arc center, led to values of interplanar spacings several percent greater than the correct values. It was found possible to reduce the values determined from each pattern to agreement with the correct values simply by adding to each measured radius a fixed amount. The amount to be added varied from pattern to pattern, its range being from 0.2 mm to about 1.2 mm.

It appeared, therefore, that each pattern as a whole had been shifted toward the base line upon which the center of the arcs should fall. Such a shift can reasonably be explained only by the assumption that the specimen had acquired a positive charge under the influence of the electron beam. As before mentioned, in all cases except two, the direction of the shift has been such as to indicate a positive charge.

CAUSE OF PATTERN SHIFT

It has not been found possible to demonstrate, except by the direction of pattern shift and the general direction in which the "undeviated" beam, which passes the specimen during the exposure, is spread, that the charge is positive. However, it has been shown that a positive charge applied to a specimen will produce just such a shift. A layer of zinc oxide, which normally does not produce a shift, was placed upon a metal specimen support (Fig. 3). The specimen was placed upon small pieces of glass to insulate it from the specimen holder and a wire lead attached to the specimen support was taken outside the camera. The procedure outlined below was

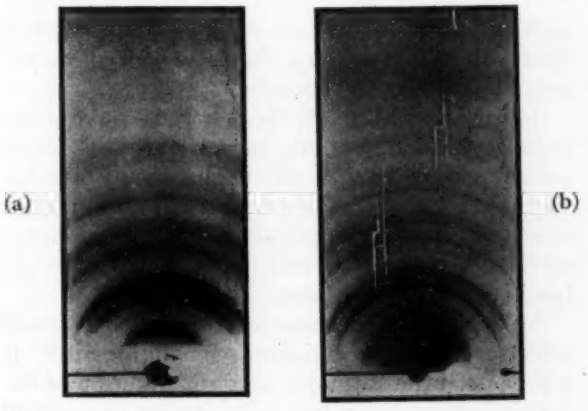


FIG. 2.

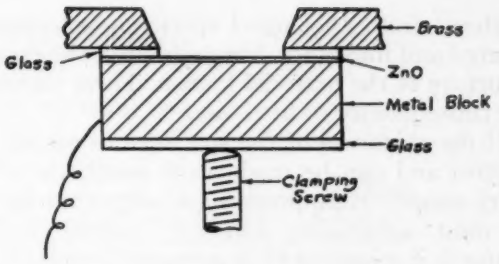


FIG. 3.

then followed. With specimen withdrawn from the electron beam, a base line was traced on the photographic plate. The specimen was then made 40 volts positive with respect to the camera, put into position for diffraction and an exposure made. No base line was traced after the exposure. Next an exposure was made on another portion of the plate in the usual way, with the specimen at zero potential. The arcs of the pattern made with the specimen positively charged were found to be shifted toward the base line so that their center was 0.8 mm below it, while those of the pattern made with the specimen uncharged were found to have their center in the base line. Shifts of 0.2 mm to 1.2 mm such as we observe would therefore be produced by charges which raised the specimen surfaces to potentials of 10 to 60 volts positive.

Such charges can originate in the circumstances only by the emission of sufficient secondary electrons from the specimen to more than counterbalance the absorption of primary electrons from the incident beam, i.e., only if the coefficient of secondary electron emission is greater than one. The published work on secondary emission produced by high voltage electrons appears to show that the coefficient is less than one for voltages above about 20-kv. The work seems, however, to have been done with the primary electron beam normal to the surface being studied, not with grazing incidence such as occurs in electron diffraction by the reflection method.

There are several courses which might be followed in dealing with the pattern shifts which are encountered. If the pattern contains one well-defined arc of fairly large radius, its center may be located by the use of dividers or a similar method. If the crystal structure of the substance which gave the pattern is known and two orders of diffraction from a certain set of planes appear, these may be used to locate the center. Unfortunately, these conditions are seldom fulfilled.

Patterns from a charged specimen are usually blurred and indistinct. Very seldom is the crystal structure of the material known. These methods are consequently of little use.

If the thickness of the specimen is subject to control and can be made very small, the difficulty usually disappears since very thin layers of most substances will not become highly charged, if mounted on a grounded metal plate. In many cases such control is not possible and the best solution of the problem is to irradiate the specimen with an auxiliary beam of low voltage electrons to prevent a charge from accumulating.

USE OF AUXILIARY ELECTRON GUN

The auxiliary beam used in this laboratory is from an electron gun set into the side of the diffraction camera at such a place that the beam is directed normally upon the specimen surface. For the basic features of the gun design, we are indebted to J. E. Ruedy of the RCA Laboratories. An accelerating potential of 400 v is applied to the gun. With this arrangement, it has been possible to reduce to a negligible value shifts of pattern from all the specimens examined since it was placed in operation, and to remove from the recorded patterns all evidence of electrical charging of the specimens. The improvement can be seen in the diffraction patterns of Fig. 2. These were made from a fairly thick layer of TiO_2. The pattern of 2(a) was made without the use of the auxiliary electron beam; that of 2(b) with it. The pattern shift, beam distortion, and the broadening of the diffraction lines are all clearly evident in 2(a); and are just as clearly absent in 2(b). The broadening of the lines probably occurs because various parts of the specimen surface became charged to different potentials. In the pattern of 2(a) it is pronounced enough to destroy all trace of the single crystal reflections such as are seen in 2(b).

Through use of the auxiliary electron beam the electrical charges which produce the pattern shift have been reduced to negligible amounts both on specimens which become positive and on the few specimens which become negatively charged. In particular, the auxiliary beam was found to reduce both positive and negative charges produced on a piece of microscope cover glass.

It was found that the sign of the charge on the glass could be changed merely by altering the inclination of the glass surface to the main electron beam. If the glass was set with its surface nearly parallel to the direction of the beam, the charge was strongly negative. As the inclination to the beam was increased the charge, as estimated from the deflection of the beam, decreased until it fell to zero and reversed in sign. Further increasing the inclination caused the positive charge to increase. At all stages of the process of changing the inclination, the deflection of the beam, and hence the charge on the specimen, could be reduced to zero by turning on the auxiliary electron beam.

A rational explanation of these results solely in terms of variations in secondary electron emission from the glass with changes of inclination of the specimen surface to both the main (40 kv) and auxiliary (400 v) electron beams does not seem possible. It does seem necessary to assume a change of the secondary emission coefficient for the 40-kv electrons from less than to greater than unity as the grazing angle of incidence on the specimen is increased; but a simultaneous change in the opposite sense of the secondary emission coefficient for the 400-v electrons produced by a slight alteration in their nearly normal direction of incidence is unlikely. It seems more reasonable to suppose that a large part of the action of the 400-v electrons in reducing charges on specimens is the result of ionization of the residual air in the camera near to the specimen surface. From the cloud of ions thus formed, a charged specimen could draw enough ions of opposite sign to reduce its charge to a small value.

SUMMARY

Distortion and shift, relative to their supposed center, of electron diffraction patterns made from poorly conducting materials by the reflection method have been observed. The direction of the shift and other evidence show these effects to be caused by a charge, usually positive, accumulated on the specimen.

The magnitudes of the charges have been reduced and their effects on the diffraction patterns have been made negligible by irradiation of the specimens by a beam of 400-v electrons directed perpendicularly to the specimen surface. Both positive and negative charges have been reduced in this way.

An explanation of these results solely in terms of secondary electron emission seems impossible. It is suggested that a major part of the action of the 400-v electron beam lies in ionization of residual gas in the neighborhood of the specimen surface.

A 100-kv Electron Microscope

L. MARTON
Division of Electron Optics, Stanford University
(Received September 30, 1944)

In this paper a transmission type electron microscope with magnetic lenses is described. The electron speed can be varied between 30 ekv and 100 ekv. The magnification of the instrument is produced in three stages. The instrument has improved air locks, hydraulically operated stage movement, and a stage tilting device up to $\pm 15\frac{1}{2}°$. It is also provided with means for bright and dark field illumination and for conversion into a diffraction camera. Improved electrical circuits provide the necessary stability of the power supplies.

THE electron microscope constructed at Stanford University differs in many respects from previous instruments.[1–9] The following description without attempting to be complete, will mostly emphasize the features which are different.

In the general design of the instrument, care was taken to make it adaptable to many modifications. In some respects the instrument can be described as a "universal" electron microscope by using the word "universal" in the same sense von Ardenne did.[6] The instrument is upright with electron gun on top and a transparent fluorescent screen at the bottom. It is supported by the vacuum manifold and is therefore accessible from all sides. The power supply forms a separate unit and so does the control desk. The accelerating voltage range is from 30 kv to 100 kv in steps of 10 kv. It is provided with improved object chamber and photographic chamber mechanisms. Its object chamber is built for use of a new type specimen holder which offers a certain number of advantages. The general appearance of the instrument and the cross section drawing are shown in Figs. 1a and 1b.

ELECTRON GUN

The construction of the electron gun is shown in Fig. 2. A ceramic bushing separates the cathode from the anode. Centering of the two electrodes is achieved by grinding the inner surface of the insulator to a conical shape into which both the anode and cathode fit. The filament holder has coaxial leads and is centered in the guard electrode by means of a machined ceramic sleeve (Alsimag 222). The beam cross over is varied by means of a bias potential applied to the guard electrode. For the purpose of tilting adjustment of the filament, the electron gun is mounted on a self-aligning type ball

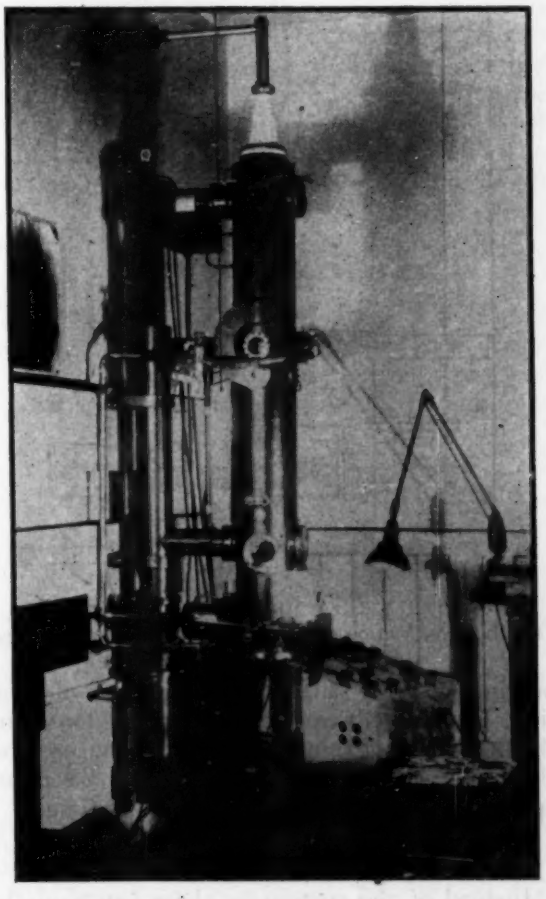


FIG. 1a. View of the Stanford electron microscope.

[1] L. Marton, Ann. et Bull. Soc. Roy. Sci. Med. Nat. Bruxelles, 92–106 (1934).
[2] E. Ruska, Zeits. f. Physik **87**, 580–602 (1934).
[3] L. Marton, Bull. Acad. Belg. **21**, 606–617 (1935).
[4] B. v. Borries and E. Ruska, Wiss. Veröff Siemens-Werken **17**, 99–106 (1938).
[5] A. Prebus and J. Hillier, Can. J. Research **17**, 49–63 (1939).
[6] M. v. Ardenne, Zeits. f. Physik **115**, 339–368 (1940).
[7] L. Marton, Phys. Rev. **58**, 57–60 (1940).
[8] V. K. Zworykin, J. Hillier, and A. W. Vance, Trans. A.I.E.E. **60**, 157–162 (1941).
[9] A. Prebus, Ohio State Univ. Eng. Exp. Sta. News. **14**, 6–32 (1942).

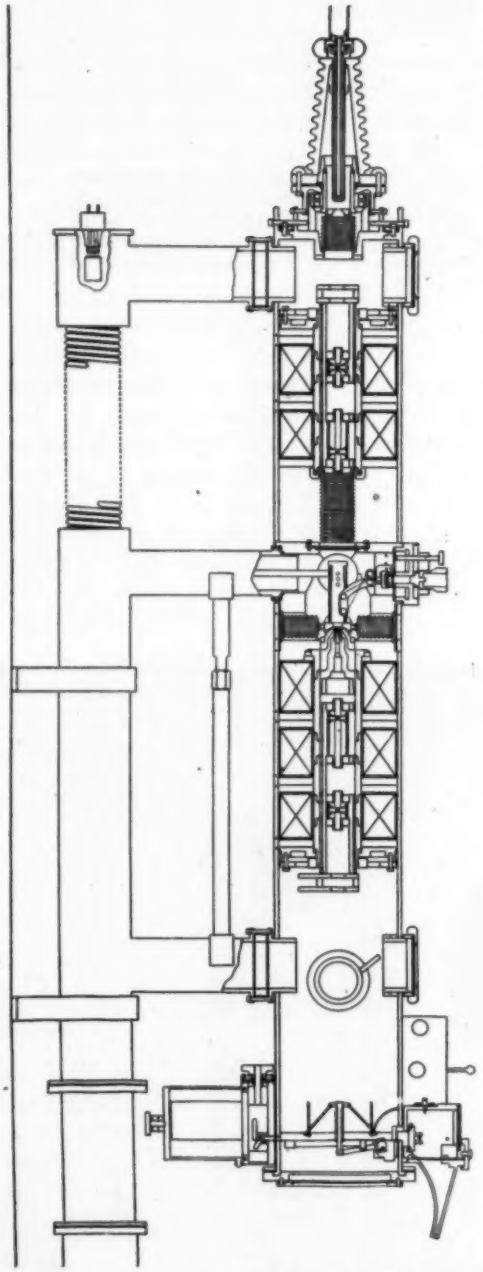


FIG. 1b. Cross section of the Stanford electron microscope.

bearing with the cathode in the center of rotation. The filament can be taken out by loosening the upper gasket tightening nut. For purposes of cleaning, the whole gun can easily be taken apart and reassembled in a very short time.

OPTICAL ELEMENTS

Instead of the customary three lenses of the compound microscope this instrument is built with five lenses, two of which belong to the condenser system and three to the imaging system. The two lenses of the condenser system are constructed so that they can produce a reduced size image of the source on the specimen (for the purpose of electron diffraction, etc.). The three lenses of the imaging system (commonly called in this laboratory: objective, intermediate, and projection lenses) are for the purpose of producing a wide range of magnifications by filling out the visual field evenly at relatively low magnifications. For the purpose of the alignment of the instrument the position of the objective lens was selected as a reference point.* The alignment of the lenses is done by shifting the pole pieces of the individual coils instead of shifting the whole coil. Description of

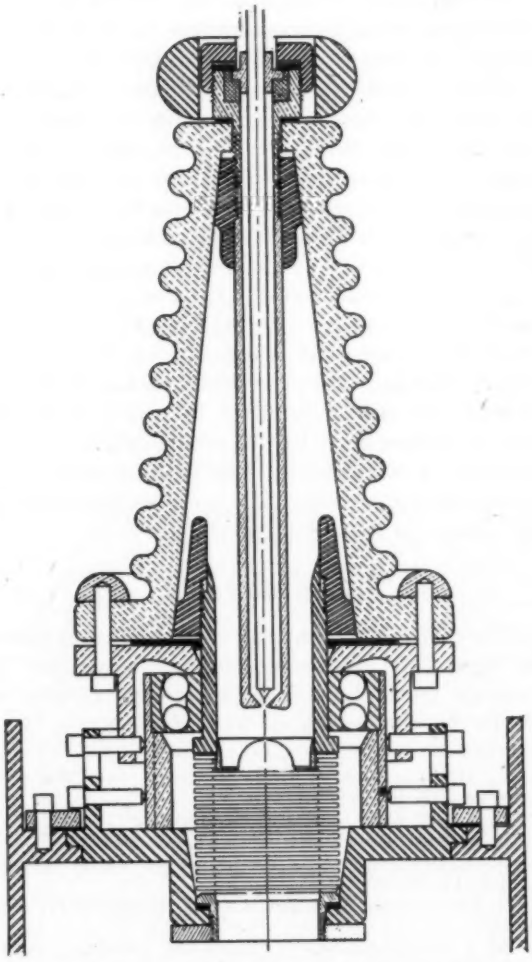


FIG. 2. Cross section of the electron gun.

* In most instruments which are supported from below, the projection lens is chosen as fixed reference point.

this type of alignment is given below for the two condenser coils only, as the intermediate and projection coil unit is almost identical in its construction to the design of the two condenser coils.

CONDENSER COILS

The vacuum enclosure consists of an iron tubing with brass spacers (shown in Fig. 3), and is therefore hard soldered. Two coils, made of 8000 turns of No. 23 wire each, can be slipped over the outer surface of this composite tube as well as the steel slide flanges. An outer shell of steel tubing, fitting over the flanges and the coils, completes the outer magnetic circuit. The advantage of this design is that the coil can be taken out separately for inspection and replacement.

The four pole pieces of the two lenses are held together by a brass sleeve. They slide with very little clearance between corresponding iron disks. The lateral movement can be achieved by means of eccentric rods operated through the gear arrangement shown on Fig. 4. To avoid asymmetry of the field, when the pole pieces are in an off-set position, the pole pieces are provided with long sleeves extending beyond the flat disks.

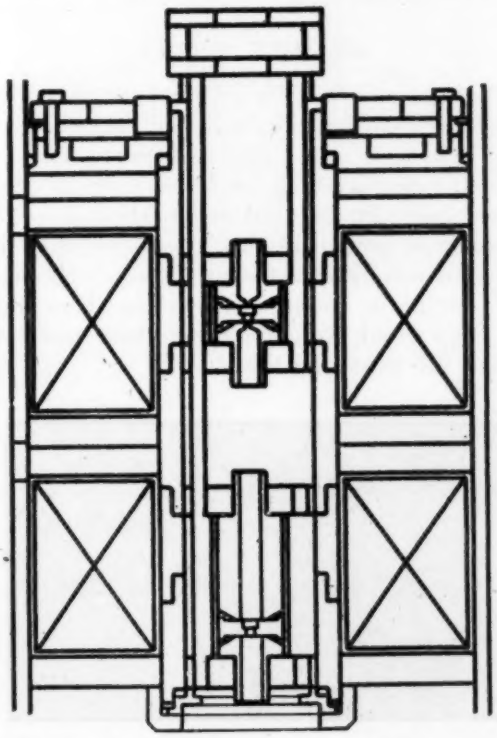


FIG. 3. Cross section of the assembly of two lenses forming a unit.

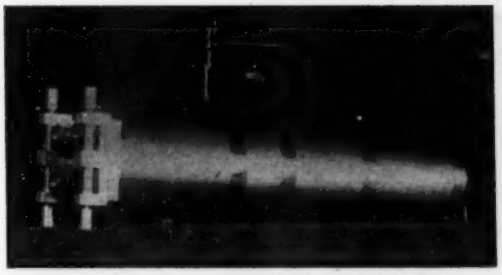


FIG. 4. View of the pole piece assembly for the lenses shown on Fig. 3.

OBJECTIVE LENS

The objective lens, provided with the same coil as all the other lenses, is distinguished from them by the fact that the object chamber housing, made of steel, forms part of the magnetic circuit. The pole pieces sit in a pole piece holder to which the stage of the microscope is fixed. The pole pieces fit into a well-machined cylindrical boring in the pole piece holder. For purposes of cleaning their aperture the pole pieces can easily be pulled out through one of the lateral openings of the object chamber. A series of different pole pieces have been provided for the objective lens, the one having the best characteristics being represented in Fig. 5. The aperture of the objective lens is placed approximately 10 mm below the pole piece gap.[10]

OBJECT CHAMBER

The two main parts of the object chamber are the air lock and the stage. The air lock design follows to some extent the mechanism developed by Mahl. Its volume is reduced to a minimum

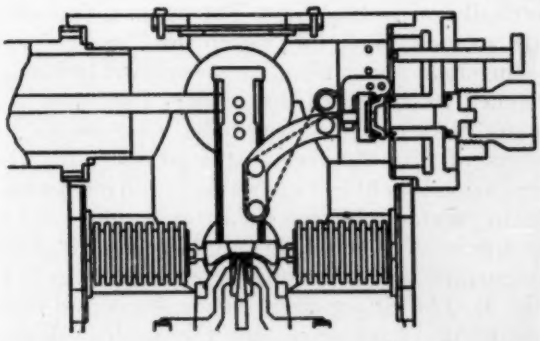


FIG. 5. Cross section of the object chamber showing also the pole pieces of the objective lens.

[10] L. Marton and R. G. E. Hutter, Phys. Rev. **65**, 161–167 (1944).

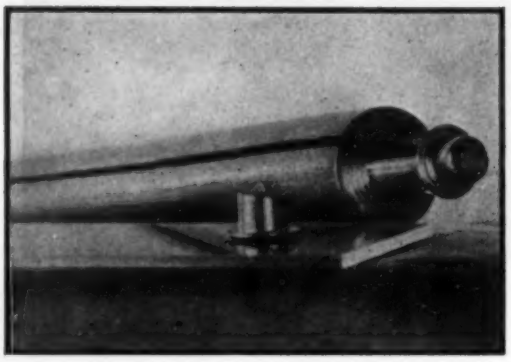


FIG. 6. Specimen holder and tong (one specimen holder placed on a standard glass microscope slide).

by mounting the specimen holder in a small bucket-like support, the flanges of which are clamped in the closed position in between two "gates" (Fig. 5). In that position the specimen holder can be removed through the front opening by means of a special tong shown on Fig. 6. This tong fits with relatively little clearance in the guide hole of the air lock and makes it, therefore, very easy to introduce and withdraw the specimen. The total time required to transfer a specimen from the stage of the microscope in the hands of the operator is 5–8 seconds and the introduction of a specimen into the microscope stage is 20–30 seconds including the pumping time of the air lock. The front gate is of the breech-lock type. The specimen holders themselves will be described further below. The "bucket" is held by two endless chains operated by means of a gear mechanism from the outside.* When the inner gate is open the "bucket" can be moved first in a horizontal position toward the optical axis of the instrument and then lowered vertically on to the stage. The stage is provided with a tapered hole into which the "bucket" fits.

The stage mechanism, as mentioned before, is fastened to the pole piece holder. The stage is a square piece of brass, suspended from above by means of four relatively stiff brass rods. The suspension is held by two brackets, hinged around bearing surfaces having their axis at the level of the specimen. The whole stage can be tilted for the purpose of stereoscopy by as much as $\pm 15\frac{1}{2}°$ (Fig. 5). The tilting angle can be measured from outside the microscope by the optical device shown on Fig. 7. It consists of a viewing telescope with a cross wire and a long index, turning together in the telescope mount. In viewing the upper part of the stage supporting bracket, which is visible through a lateral port of the object chamber, the edge of the bracket can be made to coincide with the cross wire of the viewing telescope. The sine of its tilting angle can then be read on the scale attached to the telescope mount. The handle seen in the background of Fig. 7 is the one operating the tilting mechanism.

The stage movement, instead of being mechanical, is hydraulic. Four metal bellows, with conveniently shaped shoes, press against the sides of the stage. The bellows are filled with a liquid and are connected through capillary tubing to another set of four bellows locked together and commonly displaced by a joy stick (Fig. 8). Four more smaller bellows are operated in pairs by means of micrometer screws for the vernier adjustment of the stage. The complete stage control assembly is placed in the immediate vicinity of the main viewing port, thus permitting easy scanning of the specimen. After one year of operation of this stage movement it is still entirely free of any jerkiness and permits a smooth scanning of the specimen even at the highest magnifications.

SPECIMEN HOLDER

The specimen holders adopted for this microscope are shown on Fig. 6. They consist of wire mesh which is rolled flat and then shaped as indicated in Fig. 9. A plane surface, of about 1 mm in diameter and perpendicular to the optical axis, constitutes the observed part. It is joined by a conical surface of 45° to a cylindrical part, fitting in commercial eyelets. The punch and dies used for preparing these specimen holders are shown on Fig. 10.

FIG. 7. Viewing telescope for measuring the stereoscopic angle.

* This and several other features were suggested and developed by B. F. Bubb.

Specimen holders of this design have been in operation here for over a year. They have the following advantages: 1. They are easy to handle. 2. The specimens can be easily prepared at a well-defined part of the specimen holder. 3. They are "polarized," i.e., the surfaces which support the collodium film or other films cannot be confounded with the "wrong" side of the specimen holder. 4. They can be easily filed away for further use. 5. By means of the special tong described above, their insertion and withdrawal to and from the microscope needs no particular skill on the part of the operator.

PHOTOGRAPHIC CHAMBER

The photographic chamber is of the magazine type, carrying a maximum of 24 plates, $3\frac{1}{4}'' \times 4''$

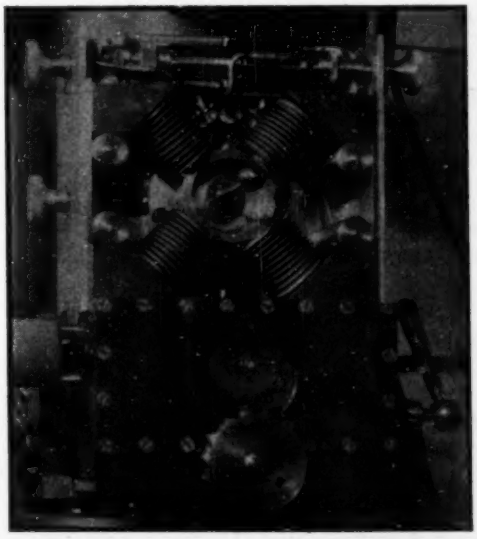


FIG. 8. Stage adjusting mechanism (with its cover off) placed on top of the receiving end of the photographic chamber.

size, from which the exposed plates can be individually sluiced out in the atmosphere after exposure (see lower part of Fig. 1). It constitutes an improvement on the type described by Brüche,[11] inasmuch as the magazine itself can be separated from the microscope by means of an air-tight gate. Further improvement consists in a movement of the plate clear across the microscope, thus providing the means for partial exposures for any part of the plate.

[11] E. Brüche and E. Gölz, Jahrbuch d. AEG-Forschung 7, 60–66 (1940).

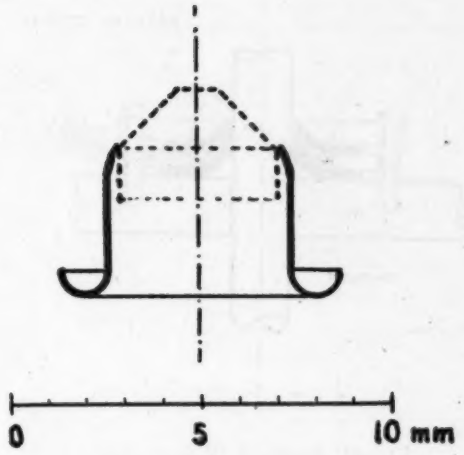


FIG. 9. Cross section of specimen holder.

From the plates stacked up in the magazine the lowest is pulled out for exposure by means of a catcher. This catcher consists of a flexible steel tape (taken from a measuring tape) provided with a hook on its outer end and operated from a drum. The plateholder slides along on two rails. After exposure is completed it can pass through a second gate in the receiving part of the chamber where it drops into a "chute." After closing the gate of the receiving part, the "chute" can be detached and carried to the darkroom, without interruption of the operation of the microscope.

The screen, which is, as usual, the photographic shutter, is made in two halves opened symmetrically in the center. Partial exposures can be made in a center position without the use of added masking devices. With the fluorescent screen open and with no photographic plate in the central position the photographic chamber mech-

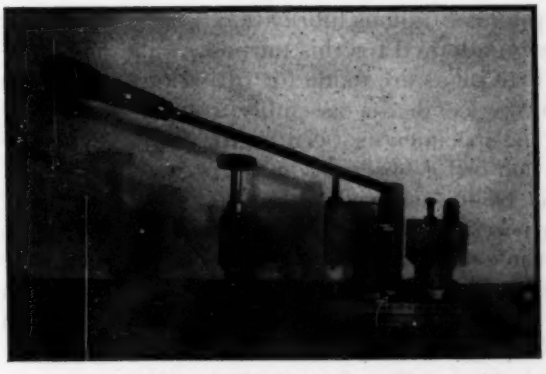


FIG. 10. Punch and die used for preparing the specimen holders.

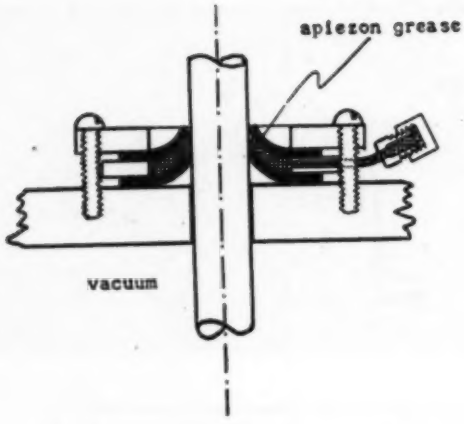


FIG. 11. Modified "Wilson-seal."

anism does not at all obstruct the view of a second fluorescent screen closing up the bottom of the instrument. The image on the fluorescent screen can be viewed simultaneously by three observers through three portholes provided with mirrors at approximately 45° angle. The portholes are large enough for binocular vision (3″ free diameter). The bottom fluorescent screen is demountable and can be replaced by any attachments to the microscope, such as an electron microspectroscope.[12]

GENERAL ASSEMBLY

Almost everywhere Ameripol gaskets are used for maintaining a good vacuum. Mechanical controls, such as lens alignment, air locks, photographic plate movement, fluorescent screen, etc. are operated through "rubber glands." These "rubber glands" (now often called Wilson-seals)[13] differ to some extent from the ones usually described in that they are provided with a grease cup for continuous lubrication. Fig. 11 shows the design adopted for this purpose.

Provisions are made for tilting of the electron source—condenser assembly, round the position of the specimen as a center, for dark field illumination or for observation at grazing incidence of the beam as used before for metallographic purposes by Borries and Ruska.[14] Being built as a three-stage microscope its final image can be observed at low magnification and a great width of field, thus making unnecessary the use of an intermediate fluorescent screen and its viewing ports.

ELECTRICAL EQUIPMENT

The power supplies, for both the high voltage and the coil currents are degenerative type feed-back regulators.[15,16] The high voltage equipment, as mentioned before, can be operated between 30 kv and 100 kv with a constancy of one part in 20,000. The bias potential of the guard electrode is obtained from a potential divider which is part of the main feed-back bleeder resistor.

The coil current supplies are modifications of similar ones described before. On Fig. 12 is shown the modified circuit as compared to one of the well-known circuit designs.[17] It offers the advantage of maintaining the regulator much more constant for varying loads of the power supply by reducing the relative charge of the screen potential of the control tube. Fig. 13

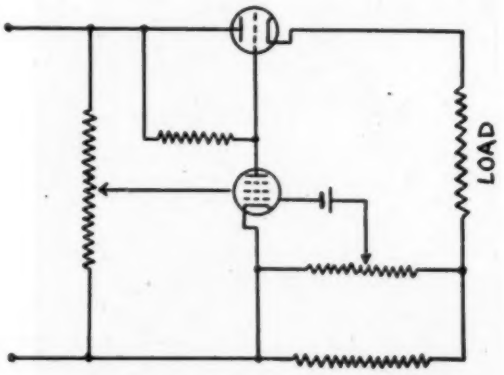


FIG. 12a. Current regulator.

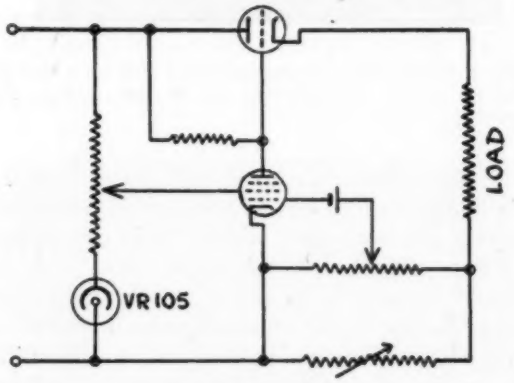


FIG. 12b. Modified circuit.

[12] L. Marton, Phys. Rev. **66**, 159(A) (1944).
[13] R. R. Wilson, Rev. Sci. Inst. **12**, 91–93 (1941).
[14] B. v. Borries and S. Janzen, V.D.I. **85**, 207–211 (1941).
[15] H. S. Black, Bell Sys. Tech. J. **13**, 1–18 (1934).
[16] A. W. Vance, RCA Rev. **5**, 293–300 (1941).
[17] L. Marton and R. G. E. Hutter, Phys. Rev. **61**, 205–206 (1942).

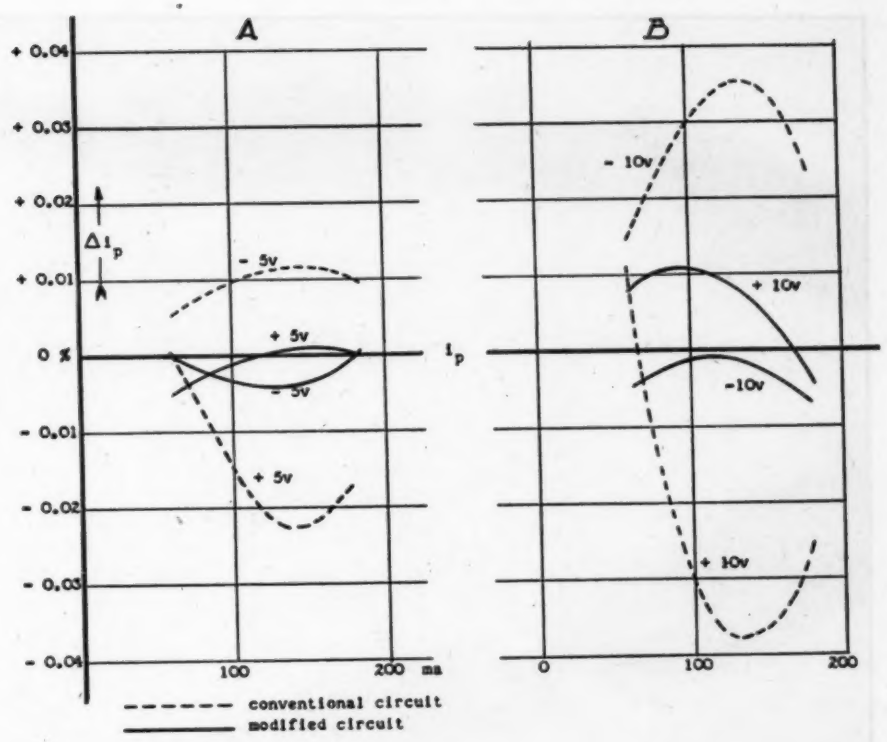


FIG. 13. Changes in percents of the output current in function of the total output, for a variation of the input voltage of (A) ±5 volts, (B) ±10 volts.

shows a comparison of the two circuits for various output currents and for line voltage variations of ±5 volts and ±10 volts.

FIG. 14. Magnesium oxide (extra light magnesia).

Control of the different power supplies is mounted on a small control desk (Fig. 1a). The control desk contains also switches for vacuum pumps and gauge circuits for measuring the vacuum. The control desk is mounted on casters and can be easily moved to the most convenient position for the operation of the instrument.

PERFORMANCE

The performance of the instrument can be judged from the micrographs, Figs. 14 and 15. Results of wide-angle stereoscopy were reported earlier in this journal.[18] Briefly reported were also our findings on poliomyelitis virus, a notoriously difficult subject.[19]

Many friends and colleagues contributed to the completion of this instrument. Most of the mechanical work was expertly done by Mr. B. F. Bubb sometimes with the help of J. Grebmeier, A. Jansse, and F. Pool. The electrical circuits were made and improved by Dr. R. G. E. Hutter with part-time help of Dr. S. W. Grinnell. Professor F. Molnar helped prepare drawings. Calibration of the instrument was to a great extent carried out by Mr. H. L. Snyder.

[18] L. Marton, J. App. Phys. 15, 726 (1944).
[19] H. S. Loring, C. E. Schwerdt and L. Marton, Phys. Rev. 65, 354 (1944).

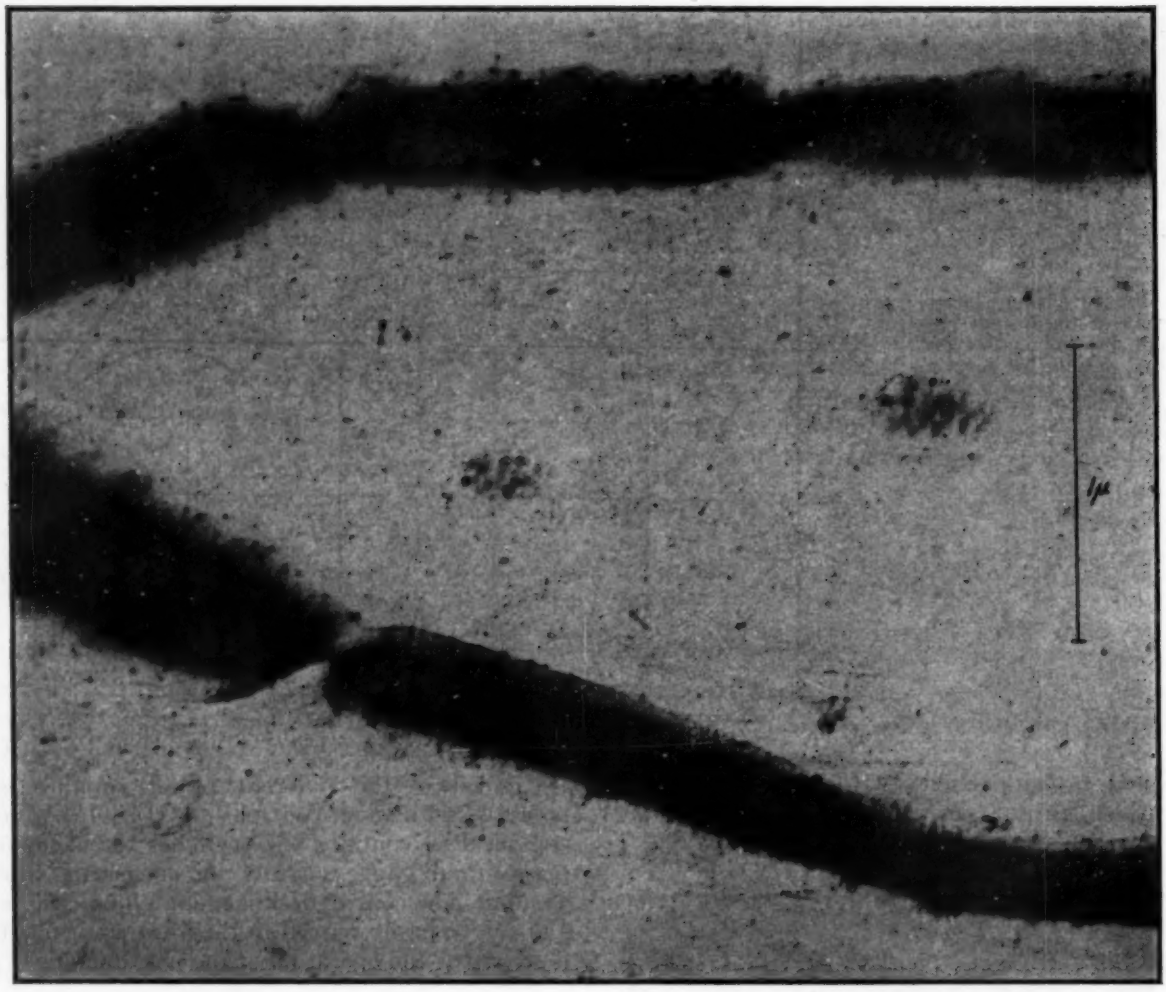


FIG. 15. Vibrio comma (specimen courtesy of Professor E. W. Schultz).

The writer also takes pleasure in acknowledging help and advice of several of his colleagues of the departments of physics and of electrical engineering. And, last but not least, he acknowledges the constant help and efforts of his wife, Mrs. C. Marton.

Solution of the Diffusion Equation Applicable to the Edgewise Growth of Pearlite

W. H. BRANDT
Westinghouse Electric and Manufacturing Company, East Pittsburgh, Pennsylvania

(Received October 6, 1944)

The diffusion equation is transformed to a set of coordinates moving with the pearlite interface and a solution applicable to the problem obtained in the form of an infinite series of terms. Using the first three terms, the edgewise velocity of pearlite growth is calculated for a plain carbon eutectoid steel using data most of which are obtained by extrapolation. The values obtained show reasonable agreement with values for the rate of pearlite nodule growth determined by Hull, Colton, and Mehl. The velocity increases with decreasing temperature, as expected, and it is shown that this is caused by the change in the solubilities of ferrite and cementite in austenite with temperature. The theory predicts curved ferrite-austenite and cementite-austenite interfaces and the carbon concentration in austenite is shown to vary across each of these interfaces.

INTRODUCTION

A STEEL of a composition and temperature lying between the lines *GS* and *ES* on the equilibrium diagram (Fig. 1) consists of austenite

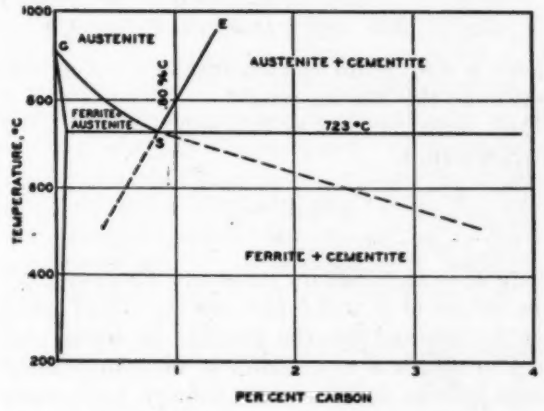


FIG. 1. Iron carbon constitution diagram (*Metals Handbook*, 1939 edition, p. 368). Extrapolations of ferrite and cementite solubilities are shown in dotted lines.

(face-centered cubic iron) with the carbon completely in solution. To the right of the line *ES*, cementite (Fe_3C) is formed; to the left of the line *GS*, ferrite (body-centered cubic iron) is formed. The line *ES* is known as the cementite solubility line, the line *GS* is known as the ferrite solubility line and the temperature and carbon concentration at which they meet are known as the eutectoid temperature and eutectoid carbon concentration, respectively.

Below the eutectoid temperature the steel may be to the right of the cementite solubility line and to the left of the ferrite solubility line simultaneously, and a constituent known as pearlite consisting of alternate layers of cementite and ferrite is formed provided that the temperature of formation is not below about 600°C. The hardness of the products of austenite decomposition becomes progressively greater as the temperature of formation decreases. The formation of pearlite is of considerable importance in the hardening of steel for the metal transformed into pearlite does not transform into harder constituents, and therefore limits the hardenability.

The growth of pearlite has been studied in detail, largely by Mehl and his collaborators.[1–4] It has been shown that pearlite grows in nodules of roughly spherical shape, each nodule consisting of colonies. A colony is a volume of pearlite of substantially uniform lamellar and crystallographic orientation and is the largest unit of pearlite formed by continuous growth. The rate of formation increases sharply with decreasing temperature below the eutectoid temperature and depends on two factors, nucleation rate and growth rate of nodules. Both have been measured. It has been shown that nucleation rate is affected by austenite grain size and austenite heterogeneity while growth rate is not so affected.

It has been pointed out by Mehl[2] that the growth rate G will increase with increasing coefficient of diffusion and will increase with decreasing pearlite spacing. Hultgren[3] has suggested that the rate of growth would be affected by the cementite and ferrite solubilities and that these could be determined by extrapolation below the eutectoid temperature, but there have been

[1] F. C. Hull and R. F. Mehl, Trans. Am. Soc. Metals **30**, 381 (1942).
[2] R. F. Mehl, Physics of hardenability. Hardenability of alloy steels, Am. Soc. Metals Symposium (1939).
[3] Pellissier, Hawkes, Johnson, and Mehl, Trans. Am. Soc. Metals **30**, 1049 (1942).
[4] Hull, Colton, and Mehl, Trans. A. I. M. E. **150**, 185 (1942).

no relations worked out connecting the various quantities involved.

It is furthermore still not clear why pearlite has a lamellar structure. A mechanism was suggested by F. C. Hull and R. F. Mehl[1] whereby a lamellar structure could be started, but no justification was given for the fact that it would continue to grow with uniform lamellar spacing. Furthermore, no explanation has been given as yet of the factors which control the spacing during edgewise growth though the spacing has been measured as a function of temperature and some information is available on the effect of various alloying additions on the spacing.[3]

In surveying these problems it is clear that the next step in the theory of pearlite formation should be a solution of the diffusion equation applicable to pearlite growth. It will be seen that this contributes to our knowledge of growth rates and indicates the existence of lamellar structure. It suggests a mechanism that may control the spacing, but the latter point is as yet uncertain.

THE DIFFUSION EQUATION

The equation to be solved is:[2]

$$\frac{\partial^2 C}{\partial x'^2}+\frac{\partial^2 C}{\partial y'^2}=\frac{1}{D}\frac{\partial C}{\partial t}, \tag{1}$$

where C is the carbon concentration, x' and y' are ordinary rectangular coordinates, t is the time, and D is the diffusion coefficient. This equation is identical with the heat conduction equation, and solutions applicable to many of the simpler stationary boundary cases are well known. Obviously these solutions will not be helpful in this case since in this problem the boundary moves.

It is clear, however, that the growth of pearlite proceeds at a uniform velocity once growth has been well started. This is a logical assumption to make by way of distinguishing pearlite growth on one hand from such phenomena as precipitation hardening in which growth proceeds with a velocity which decreases with time, and on the other hand, from phenomena such as simple transformation of austenite to ferrite in the absence of carbon which appears to have no limitations on its velocity and occurs very rapidly. The velocity of nodule growth has been measured furthermore by Hull, Colton, and Mehl[4] and shown to be constant over virtually the entire period of growth of a pearlite nodule. It seems likely that edgewise growth of pearlite will behave in a similar manner. If the velocity is constant, the carbon concentration must also be constant provided that it is measured with reference to the pearlite interface. In making this assumption it has been assumed also that growth of both ferrite and cementite is continuous and not stepwise. For the sake of simplification, the analysis is restricted to the case in which the interface is at right angles to the x axis and the direction of growth.

We dispose then of the moving boundary by transforming to a set of axes stationary with respect to the interface, that is, moving with a velocity, V, in the x direction so that $x=x'-Vt$ and $y=y'$. Then since

$$\frac{\partial C}{\partial t}=\frac{\partial C}{\partial x}\frac{\partial x}{\partial t}=-V\frac{\partial C}{\partial x},$$

Eq. (1) becomes,

$$\frac{\partial^2 C}{\partial x^2}+\frac{\partial^2 C}{\partial y^2}+\frac{V}{D}\frac{\partial C}{\partial x}=0, \tag{2}$$

where x and y are the coordinates taken with respect to the moving axes.*

Any term $Ke^{ax}\cos by$ is a solution of Eq. (2) provided that,

$$a^2+a\frac{V}{D}=b^2, \tag{3}$$

where K is an arbitrary constant. However, certain values of a and b give rise to terms which may be ignored for the pearlite problem. Any term in which b is imaginary, or complex, becomes infinite as y becomes infinite and cannot correspond to an interface at right angles to the x axis. If b is not imaginary or complex, a cannot be complex. Negative values of b give terms indistinguishable from those arising from positive values and may be ignored so that b is either zero or real and positive. For each such value of b except $b=0$, there is one negative and one positive value of a. The positive values of a give rise to terms which become infinite as x becomes infinite and may be ignored.

Applying the usual solution for a quadratic to Eq. (3),

$$a=\frac{-\frac{V}{D}-\left\{\left(\frac{V}{D}\right)^2+4b^2\right\}^{\frac{1}{2}}}{2}, \tag{4}$$

* This equation was previously suggested for the pearlite problem in a thesis submitted by F. C. Hull to the Graduate Committee of the Carnegie Institute of Technology, in partial fulfillment of the requirements for the Degree of Doctor of Science, June, 1941.

ignoring the positive root. When $b'=0$,

$$a=0 \quad \text{or} \quad -\frac{V}{D}.$$

Since the sum of any two solutions of Eq. (2) is also a solution, one writes

$$C=K'+\sum_{n=0}^{n=\infty} K_n \exp[a_n x] \cos b_n y, \qquad (5)$$

where

$$b_0=0, \quad \text{and} \quad a_0=-\frac{V}{D},$$

K', K_0, K_1, $\cdots$, K_n, $\cdots$ are arbitrary constants, a_n is always negative and b_n is positive or zero. K' may be evaluated at once since it is the only term not zero when $x=\infty$ and is, therefore, equal to C_a, where C_a is the original carbon concentration of austenite.

Equation (5) suggests a function periodic in y which in turn suggests a lamellar structure. At any fixed value of x, the equation becomes a Fourier series if we set $b_n=nb_1$ and will fit any periodic function if b_1 is chosen to correspond to the periodicity. If S_0 is the spacing of pearlite $b_1=2\pi/S_0$ since $b_1 y=2\pi$, when $y=S_0$.

Lacking detailed knowledge of the boundary, one cannot use Eq. (5) in that form. Using only the first three terms, Eq. (5) becomes

$$C-C_a=K_0 \exp[a_0 x]+K_1 \exp[a_1 x] \cos b_1 y, \quad (7)$$

where

$$a_0=\frac{-V}{D}, \qquad (8)$$

$$a_1=\frac{-\frac{V}{D}-\left\{\left(\frac{V}{D}\right)^2+\left(\frac{4\pi}{S_0}\right)^2\right\}^{\frac{1}{2}}}{2}. \qquad (9)$$

It will be seen in the next section that one has just enough knowledge of conditions at the boundary to require and determine two arbitrary constants in Eq. (7).

THE BOUNDARY CONDITIONS

It is clear that the interface carbon concentration in austenite must vary across either the cementite-austenite interface, the ferrite-austenite interface or both since there can be no discontinuity in the carbon concentration at the point where the two interfaces join. Since one has no idea what the variation is, values can only be assigned to the carbon concentration at one point of each interface, and the center is chosen for this purpose.

Let C_{af} be the carbon concentration in austenite at the center of the ferrite interface and C_{ac} the concentration in austenite at the center of the cementite interface. Then, if the origin of coordinates is located at the center of the cementite interface, one may write

$$C_{ac}-C_a=K_0+K_1, \qquad (10)$$

$$C_{af}-C_a=K_0 \exp[a_0 x_f]-K_1 \exp[a_1 x_f], \quad (11)$$

where x_f is the value of x at the center of the ferrite interface. K_0 and K_1 cannot yet be determined since x_f is also unknown.

There is the further condition that the velocity of growth must be that permitted by the rate at which carbon leaves the ferrite and reaches the cementite. This is examined in more detail in Appendix I. Equations (36) and (37) of the Appendix may be written for the rates of growth of cementite and ferrite respectively at the interface centers,

$$\frac{V}{D}=\frac{a_0K_0+a_1K_1}{C_c-C_{ac}}, \qquad (12)$$

$$\frac{V}{D}=\frac{a_0K_0 \exp[a_0x_f]-a_1K_1 \exp[a_1x_f]}{C_f-C_{af}}, \qquad (13)$$

where C_c and C_f are the carbon concentrations in cementite and ferrite, respectively.

SOLUTION FOR THE UNKNOWNS

There are now the six equations ((8) to (13)) inclusive, to determine the unknowns K_0, K_1, a_0, a_1, V and x_f.

Combining Eqs. (10) and (12)

$$\frac{V}{D}=\frac{a_0(C_{ac}-C_a)+(a_1-a_0)K_1}{C_c-C_{ac}}, \qquad (14)$$

and combining (11) and (13)

$$\frac{V}{D}=\frac{a_0(C_{af}-C_a)-(a_1-a_0)K_1 \exp[a_1x_f]}{C_f-C_{af}}. \qquad (15)$$

Using Eq. (8) and combining (14) and (15)

$$\frac{V}{D}=\frac{(a_1-a_0)K_1}{C_c-C_a}=\frac{-(a_1-a_0)K_1 \exp[a_1x_f]}{C_f-C_a}, \qquad (16)$$

which simplifies to

$$\exp[a_1 x_f] = \frac{C_a - C_f}{C_c - C_a}, \tag{17}$$

$$x_f = \frac{1}{a_1} \log_e \frac{C_a - C_f}{C_c - C_a}. \tag{18}$$

Using Eqs. (17) and (18) one gets from (10) and (11)

$$K_1 = -\frac{(C_{af} - C_a) + (C_a - C_{ac}) \exp[a_0 x_f]}{\exp[a_0 x_f] + \exp[a_1 x_f]}$$

$$= -(C_a - C_{ac}) \frac{\dfrac{C_{af} - C_a}{C_a - C_{ac}} + \left(\dfrac{C_a - C_f}{C_c - C_a}\right)^{a_0/a_1}}{\left(\dfrac{C_a - C_f}{C_c - C_a}\right)^{a_0/a_1} + \dfrac{C_a - C_f}{C_c - C_a}}. \tag{19}$$

Substituting Eq. (19) in the first part of Eq. (16),

$$\frac{V}{D} = -(a_1 - a_0) \frac{C_a - C_{ac}}{C_c - C_a} \times \frac{\dfrac{C_{af} - C_a}{C_c - C_a} + \left(\dfrac{C_a - C_f}{C_c - C_a}\right)^{a_0/a_1}}{\left(\dfrac{C_a - C_f}{C_c - C_a}\right)^{a_0/a_1} + \dfrac{C_a - C_f}{C_c - C_a}}. \tag{20}$$

Now substituting Eqs. (8) and (9) and setting

$$P = \left\{1 + \left(\frac{4\pi}{S_0} \Big/ \frac{V}{D}\right)^2\right\}^{\frac{1}{2}}, \tag{21}$$

$$P = 1 + N \frac{L^{2/(1+P)} + L}{M + L^{2/(1+P)}}, \tag{22}$$

where

$$N = 2\frac{C_c - C_a}{C_a - C_{ac}}, \quad L = \frac{C_a - C_f}{C_c - C_a}, \quad M = \frac{C_{af} - C_a}{C_a - C_{ac}}. \tag{23}$$

When N, M, and L are known, it is possible to calculate P by successive approximations. We note by inspection that the solution of Eq. (22) will occur at P greater than 1. A convenient approximate value is $1+(N/M)$. Substituting the corresponding value in the right-hand side of (22) gives an improved value for P. In most cases about four calculations are sufficient to determine the value of P to an accuracy of one percent. The value of V can then be calculated using an equation arising from Eq. (21),

$$V = 4\pi \frac{D}{S_0} (P^2 - 1)^{-\frac{1}{2}} = 4\pi \frac{D}{S_0} V_e \tag{24}$$

where

$$V_e = (P^2 - 1)^{-\frac{1}{2}}.*$$

Knowing the value of P, one can calculate K_1 and K_0 since from Eqs. (8), (9), and (16),

$$K_1 = (C_c - C_a) \frac{2}{1 + P}, \tag{25}$$

K_0 may be easily calculated using Eq. (10).

CALCULATION OF THE VELOCITY

A calculation of the velocity has been made with the best data that were found for a simple carbon eutectoid steel and the data are shown in Tables I and II.

TABLE I. Data used in calculation of P, K_0, and K_1.

T(°C)	N	L	M	P	K_0	K_1
550	34.0	0.131	6.50	3.70	4.00	−4.35
600	47.8	0.131	6.50	5.52	2.30	−2.55
650	80.5	0.130	6.50	10.25	1.12	−1.27
700	256	0.130	6.50	36.6	0.285	−0.331
720	196×10	0.130	6.50	293	0.342×10^{-1}	-0.402×10^{-1}

TABLE II. Data used in calculation of V and x_f/S_0.

T(°C)	Log V_e	Log D	Log S_0	Log $\frac{D}{S_0}$	Log V	$\frac{x_f}{S_0}$
550	9.44 −10	0.65 −10	4.60 −10	6.05 −10	6.59 −10	0.25
600	9.27 −10	1.14 −10	5.00 −10	6.14 −10	6.50 −10	0.27
650	8.99 −10	1.55 −10	5.40 −10	6.15 −10	6.24 −10	0.23
700	8.44 −10	1.93 −10	5.60 −10	6.13 −10	5.67 −10	0.31
720	7.53 −10	2.08 −10	5.96 −10	6.12 −10	4.75 −10	0.32

The values of N, M, and L were determined from the iron carbon diagram,[5] and are given in Table I. The values of C_{af} and C_{ac} are determined by extrapolation of the ferrite and cementite solubilities to the range below the eutectoid temperature** as suggested by Hultgren.[2] It is

* V_e may be thought of as containing all data affecting V that one obtains from equilibrium relations.

[5] *Metals Handbook* (Am. Soc. Metals, 1939 edition), 368.

** At this point the assumption is made that the carbon concentration at an interface during growth of the interface is the same as that concentration which is in equilibrium at the same temperature under static conditions. It has been pointed out by Hull and Mehl[1] that this is not precisely true. It seems unlikely, however, that the difference can be large enough to affect the results of these calculations appreciably.

probable that the equilibrium carbon concentrations are affected somewhat by the curvature of the interface. By choosing the center points of each interface, however, serious errors from this cause probably have been avoided. The extrapolation itself is probably the source of the greatest uncertainty in this part of the analysis. For purposes of calculation, it has been assumed that the values of $C_{\alpha f}$ and $C_{\alpha c}$ are given by the equations:

$$0.80 - C_{\alpha c} = 0.002(723 - T), \qquad (26)$$

$$C_{\alpha f} - 0.80 = 0.013(723 - T), \qquad (27)$$

where T is the temperature in degrees C. The corresponding values are plotted on the iron carbon diagram in Fig. 1. The ferrite solubility shows some curvature above the eutectoid temperature and this curvature may be continued at lower temperature. If so, the calculated velocity at lower temperature would be higher.

The values of the diffusion coefficient were obtained by extrapolation of the equation,

$$\log_{10} D = -5.00 - 6.82\left(\frac{1}{T_K} - 5.8\times10^{-4}\right)\times10^{3}, \qquad (28)$$

T_K being the absolute temperature. This equation was obtained by fitting a curve given by Cyril Wells and R. F. Mehl.[6] Theoretically it seems to be the best form of equation to use for extrapolation purposes.

The values of S_0 were calculated using

$$\log_{10} S_0 = 0.008(T - 600) + 3.00, \qquad (29)$$

S_0 being the spacing in angstroms and T being in degrees Centigrade, which was derived from a curve given by G. E. Pellissier, M. F. Hawkes, W. A. Johnson, and R. F. Mehl[3] for eutectoid steels. In this case the form of the equation is not justified on theoretical grounds but this point is not so important since very little extrapolation is required.

The use of equations makes it possible to obtain smooth curves, but it should be understood that the figures given in the constants of these equations are not significant figures and are not related to the accuracy with which the values are known. The calculated values of V are shown in Table II and in Fig. 2 which is a plot of the calculated values superimposed on the

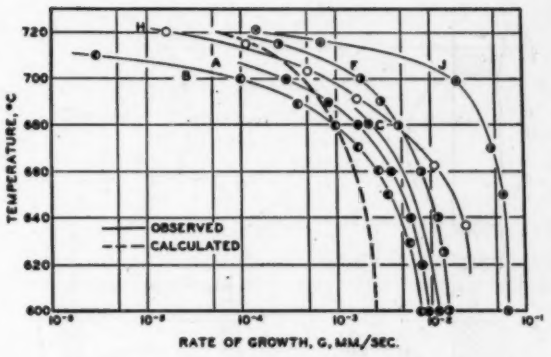


FIG. 2. Calculated values of V (dotted) superimposed on G curves of Hull, Colton, and Mehl (reference 6). J is a high purity eutectoid steel; A, B, C, H, and F are commercial steels.

curves given by F. C. Hull, R. A. Colton, and R. F. Mehl[6] for G, the rate of nodule growth. It will be seen that the shape of the curve is correct and the values are within an order of magnitude of the G values for commercial steels A and B. It is interesting that steels A and B are the same steels for which the spacings given by Eq. (29) fit best. No attempt has been made to adjust any extrapolated values so as to give a better fit to the curve.

The agreement with values for G is quite good in view of the numerous reasons why no exact agreement is to be expected: (a) Spacing data are not known for steel J which differs most from the calculated values so that agreement in this case is not to be expected. (b) The other steels contain impurities so that their equilibrium diagrams differ from Fig. 1. (c) It is not to be expected that nodule growth rate and edgewise growth rate of pearlite colonies will agree precisely* since nodule growth is an average of edgewise and sidewise growth and since there may be some delay in the starting of new colonies. (d) The pearlite interface may not always be at right angles to the growth direction as was assumed. (e) The diffusion coefficient is assumed to be constant whereas it varies with carbon concentration.[6] (f) The accumulative error caused by inaccuracy of spacing measurements and uncertainty in the extrapolations is probably in excess of a factor of two.

Table II gives values of $\log D/S_0$ used in the above calculation. It will be noted that D/S_0 remains nearly constant with temperature and that virtually all of the change in V with

[6] Cyril Wells and R. F. Mehl, Trans. A. I. M. E. **140**, 279 (1940).

* It was for this reason that the velocity is here referred to as V, whereas Hull, Colton, and Mehl use G.

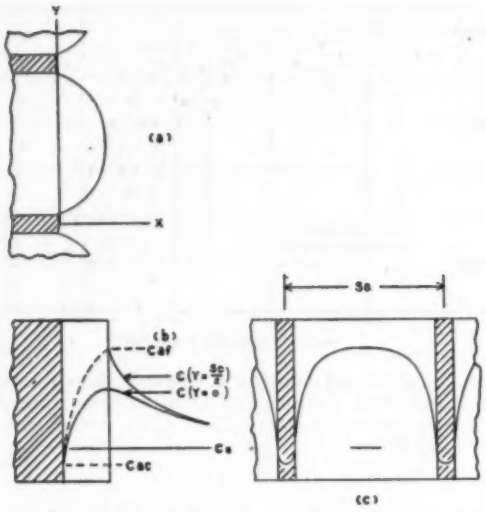


FIG. 3. Pearlite-austenite interface and carbon concentration (schematic). (a) Cross section in X-Y plane showing shape of the interface. Cross-hatched area is cementite. Curvature of cementite is exaggerated. (b) View along the Y axis showing the carbon concentration in a plane bisecting a cementite lamella ($y=0$), and in the plane bisecting a ferrite lamella ($y=S_0/2$). Intersection of the carbon concentration with the ferrite-austenite interface is shown dotted. (c) View along the X axis showing intersection of the carbon concentration with the pearlite-austenite interface.

temperature comes from V_e. In other words, nearly all of the increase in growth rate with decreasing temperature arises from the change in ferrite and cementite solubilities.

THE BOUNDARY SHAPE

The shape of the interface boundary is calculated in Appendix II and is given by:

$$\exp[a_1x_0]=\frac{b_1y_0}{\sin b_1y_0}, \tag{30}$$

and

$$\exp[a_1x_0]=\frac{C_a-C_f}{C_c-C_a}\frac{b_1\left(\frac{S_0}{2}-y_0\right)}{\sin b_1\left(\frac{S_0}{2}-y_0\right)}, \tag{31}$$

for the ferrite and cementite respectively where x_0 and y_0 are the interface coordinates. Equation (31) reduces to Eq. (17) when $y=S_0/2$.

It is an interesting prediction of this theory that the ferrite-austenite interface leads the cementite-austenite interface as given by Eq. (17). Values of x_f/S_0 calculated from Eq. (17) are shown in the right-hand column of Table I. It will be seen that the ferrite leads the cementite by a distance of about one-third of the pearlite spacing in a eutectoid steel. If present, this distance should be clearly visible on photomicrographs, but since the distance corresponds roughly to the distance grown by pearlite in a second or less, it is unlikely that the exact shape of the interface is maintained during quenching.

The shape of the interface boundary is given in Fig. 3. The carbon concentration in austenite rises along the cementite interface as the ferrite is approached and falls along the ferrite as the cementite is approached. It is not clear at the present time how this fact can be reconciled with the equilibrium relations. It is the sort of difficulty that must appear in any theory since the carbon concentration in austenite must be in equilibrium with both the cementite and ferrite at the juncture of cementite, ferrite, and austenite. While no rigorous analysis has been made, it seems likely that the factors which control these boundary conditions also control the pearlite spacing. A possible explanation is that the carbon concentration in austenite is lower at a convex ferrite surface than at a plane surface. This would be analogous to the vapor pressure at a vapor-liquid interface becoming greater as the interface becomes more convex. This should serve as a controlling factor on the spacing and in that case, one would expect that S_0 would become smaller as $C_{af}-C_{ac}$ becomes larger, as it actually does.

The carbon concentration in the neighborhood of the pearlite interface is shown in Fig. 3. This is in substantial agreement with the model suggested by F. C. Hull and R. F. Mehl[1] except for the shape of the interface.

CONCLUSIONS

Representation of carbon concentration in the neighborhood of the pearlite interface by the equation, $C-C_a \doteq K_0\exp[a_0x]+K_1\exp[a_1x]\cos b_1y$, leads to the following results: (a) Requires a lamellar pearlite structure to fit boundary conditions. (b) Yields theoretically calculated values for the rate of pearlite growth in agreement with experimental values for G to within an order of magnitude. (c) Predicts curved cementite-austenite and ferrite-austenite interfaces. (d) Agrees, except for boundary shape, with the model shown by Hull and Mehl. (e) Suggests that the determination of pearlite spacing may lie in the equilibrium between ferrite and austenite and between austenite and cementite at curved surfaces. (f) Shows that most of the increase in

growth velocity with decreasing temperature is caused by changes in the ferrite and cementite solubilities.

APPENDIX I—GROWTH OF THE INTERFACE

In Fig. 4 is shown a cementite-austenite interface which one may describe by the equation $F(x_0y_0)=0$. At the point P let the carbon gradient, grad C, be given by the vector O making an angle θ with the x axis and let the normal to the surface be N, making an angle α with the x axis. If V_n is the rate of growth parallel to N, the rate at which carbon precipitates per unit area of surface is $C_c d_c V_n$ where d_c is the density of cementite. This carbon gets to the surface in two ways: it diffuses to the surface at a rate given by $d_a D(\partial C/\partial n)$, $\partial C/\partial n$ being the normal component of the gradient, and some arrives because of boundary movement at a rate given by $d_a V_n C(x_0, y_0)$ where d_a is the density of austenite, so that

$$C_c d_c V_n = d_a D\frac{\partial C}{\partial n} + d_a V_n C(x_0, y_0). \qquad (33)$$

The densities are so nearly equal that one may write

$$V_n = \frac{D}{C_c - C(x_0, y_0)}\frac{\partial C}{\partial n}, \qquad (34)$$

$$V = \frac{V_n}{\cos\alpha} = \frac{D}{C_c - C(x_0, y_0)}\,\text{grad}\, C\frac{\cos(\theta-\alpha)}{\cos\alpha}. \qquad (35)$$

The relations

$$\text{grad}\, C = \left\{\left(\frac{\partial C}{\partial x}\right)^2 + \left(\frac{\partial C}{\partial y}\right)^2\right\}^{\frac{1}{2}}, \quad \tan\alpha = -\frac{dx_0}{dy_0},$$

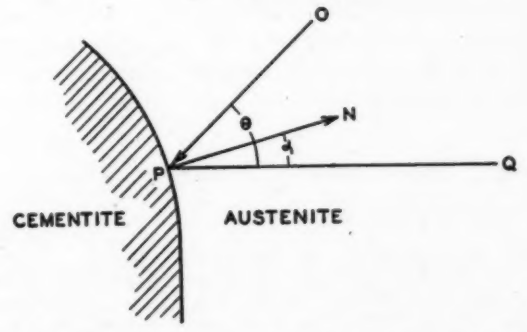


FIG. 4. Cementite interface showing normal N and vector O=grad C. PQ is parallel to X axis.

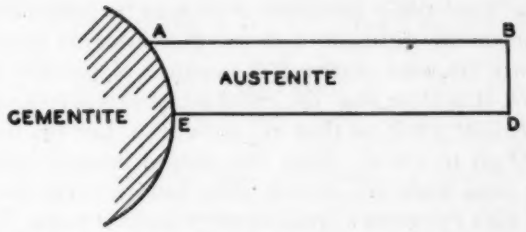


FIG. 5. Cementite interface showing volume of austenite about which one integrates to determine boundary shape. ED coincides with X axis.

$$\cos\theta = \frac{\partial C}{\partial x}\left\{\left(\frac{\partial C}{\partial x}\right)^2 + \left(\frac{\partial C}{\partial y}\right)^2\right\}^{-\frac{1}{2}},$$

$$\sin\theta = \frac{\partial C}{\partial y}\left\{\left(\frac{\partial C}{\partial x}\right)^2 + \left(\frac{\partial C}{\partial y}\right)^2\right\}^{-\frac{1}{2}},$$

reduce Eq. (35) to

$$\frac{V}{D} = \frac{1}{C_c - C(x_0, y_0)}\left\{\frac{\partial C}{\partial x} - \frac{dx_0}{dy_0}\frac{\partial C}{\partial y}\right\}. \qquad (36)$$

By the same analysis one gets for the ferrite-austenite interface,

$$\frac{V}{D} = \frac{1}{C_f - C(x_0, y_0)}\left\{\frac{\partial C}{\partial x} - \frac{dx_0}{dy_0}\frac{\partial C}{\partial y}\right\}. \qquad (37)$$

At the center of an interface $dx_0/dy_0=0$, giving Eqs. (12) and (13).

APPENDIX II—SHAPE OF THE INTERFACE

Let AE be a section of cementite-austenite interface and let $ABDE$ bound a volume in the austenite in front of the cementite with DE lying at the mid point of the cementite interface. (See Fig. 5.) AB and DE are parallel to the x axis and BD is parallel to the y axis. Let the coordinates of A be x_0 and y_0.

The net rate of flow of carbon into $ABDE$ is equal to zero so that one may write

$$d_c y_0 V C_c = \int_A^B d_a D\frac{\partial C}{\partial y}dx + \int_B^D d_a D\frac{\partial C}{\partial x}dy + \int_D^E d_a D\frac{\partial C}{\partial y}dx + \int_D^B d_a C V dy. \qquad (38)$$

The left-hand side of this equation represents the rate of carbon deposition at the interface EA,

the first three integrals represent transport of carbon by diffusion and the last integral represents transfer across BD because of motion of BD. It is clear that the third integral is zero since DE is at $y=0$, so that $\partial C/\partial y$ is zero. Let the line BD go to $x=\infty$. Then the second integral goes to zero since $\partial C/\partial x=0$. The last integral goes to $d_a C_a V y_0$ since C goes to the constant value C_a. Equation (38) then becomes

$$(d_c C_c - d_a C_a) y_0 V = \int_{x_0}^{\infty} d_a D \frac{\partial C}{\partial y} dx. \qquad (39)$$

Using Eq. (7)

$$d_a D \int_{x_0}^{\infty} \frac{\partial C}{\partial y} dx = -d_a D \int_{x_0}^{\infty} b_1 k_1 \exp[a_1 x] \sin b_1 y_0 dx$$

$$= d_a D \frac{b_1}{a_1} K_1 \exp[a_1 x_0] \sin b_1 y_0,$$

and Eq. (39) becomes, since the densities are nearly equal,

$$\frac{V}{D} = \frac{b_1}{a_1} \frac{K_1}{C_c - C_a} \exp[a_1 x_0] \frac{\sin b_1 y_0}{y_0}. \qquad (40)$$

Using the first part of Eq. (16),

$$\frac{b_1}{a_1} \frac{K_1}{C_c - C_a} \exp[a_1 x_0] \frac{\sin b_1 y_0}{y_0} = \frac{a_1 - a_0}{C_c - C_a} k_1.$$

Using Eqs. (5) and (8),

$$\exp[a_1 x_0] = \frac{b_1 y_0}{\sin b_1 y_0}. \qquad (41)$$

Equation (41) is the equation of the cementite-austenite interface. By the same procedure one may derive,

$$\exp[a_1 x_0] = \frac{C_a - C_f}{C_c - C_a} \frac{b_1}{\sin b_1} \frac{\left(\frac{S_0}{2} - y_0\right)}{\left(\frac{S_0}{2} - y_0\right)} \qquad (42)$$

for the ferrite-austenite interface. By substitution in Eqs. (36) and (37) respectively and using Eq. (7) it may be seen that the boundary everywhere grows with the velocity, V.

APPENDIX III—DEFINITION OF SYMBOLS

x', y'—coordinates with respect to stationary axes (cm); x, y—coordinates with respect to moving axes (cm); x_0, y_0—coordinates of the pearlite-austenite interface (cm); t—time (sec.); C—carbon concentration in austenite (g/g); D—diffusion coefficient of carbon in austenite (cm²/sec.); V—velocity of edgewise growth of pearlite (cm/sec.); S_0—pearlite spacing (cm except where otherwise noted); C_a—carbon concentration of cementite (g/g); C_f—carbon concentration of ferrite (g/g); C_a—carbon concentration in austenite at $x=\infty$, equals original carbon content (g/g); C_{ac}—carbon concentration in austenite at center of cementite interface (g/g); C_{af}—carbon concentration in austenite at center of ferrite interface (g/g); K, K', K_0, K_1—K_n dimensionless arbitrary constants;

$P=\{1+((4\pi/S_0)/(V/D))^2\}^{\frac{1}{2}}$;

$V_e=1/(P^2-1)^{\frac{1}{2}}$;

$N=2(C_c-C_a)/(C_a-C_{ac})$;

$M=(C_a-C_f)/(C_c-C_a)$;

$L=(C_{af}-C_a)/(C_a-C_{ac})$;

T = temperature (°C);

T_k = temperature (°K).

The Production Histories of Oil Producing Gas-Drive Reservoirs

MORRIS MUSKAT
Gulf Research and Development Company, Pittsburgh, Pennsylvania

(Received October 10, 1944)

A theory has been developed for predicting the behavior of solution gas-drive oil producing reservoirs, on the basis of previously established empirical laws on the flow of heterogeneous fluids through porous media. Treatments are given both for the simple pressure depletion history without gas injection, and those for systems in which gas is injected during the course of oil production. The specific results provided by the theoretical analysis include the ultimate oil recovery, and the pressure decline, gas-oil ratio, and productivity factor histories. Two types of gas injection have been considered, namely: (1) that in which the returned gas is supposed to diffuse through and be produced continuously with the oil zone; and (2) that in which the injected gas remains locked in the gas cap, which merely expands as oil production and gas injection proceed. In the latter case the rate of growth of the gas cap is also obtained as a function of the cumulative oil recovery.

The theory is illustrated by numerical application to a hypothetical virgin oil reservoir with an original pressure of 2500 p.s.i. producing by gas-drive, and with no initial gas cap. It is so found that if no gas is injected into the system the physical ultimate oil recovery will be 14.5 percent of the pore space, or 27.1 percent of the original stock tank oil content, assuming that the formation initially has a connate water saturation of 30 percent. The gas-oil ratio first declines as production is started, then rises sharply to a maximum of 4400 cu. ft./bbl., and finally falls steeply as the pressure is depleted to atmospheric. If there is no gas segregation and 60 percent of the produced gas is returned to the formation, the ultimate oil recovery will be increased by 27.6 percent. The gas-oil ratio history will be similar to that with no gas injection, but will rise to a maximum of 10,300 cu. ft./bbl. If 80 percent of the gas is returned, the recovery increase will be 48.8 percent, the maximum in the gas-oil ratio history reaching a value of 19,450 cu. ft./bbl. If all the gas is returned, the gas-oil ratio rise will be so rapid that by the time 20,000 cu. ft./bbl. is reached only 23.5 percent additional oil will be recovered. During these operations the productive capacities of the wells will fall by factors of the order of 10, because of the increasing viscosities of the oil and decreasing permeabilities to the oil, as the pressure declines and the oil saturation decreases.

For the case where the gas remains trapped in the gas cap, the ultimate oil recovery will be 163 percent greater than by direct pressure depletion, if the residual oil after gravity drainage is 15 percent of the pore space. This recovery will be essentially independent of the amount of gas return, although the final reservoir pressure at the time of complete gas cap expansion will be greater as more gas is returned. The increased oil viscosity and decreased permeability to the oil will here reduce the specific production capacities of the wells to $\frac{1}{4}$ or $\frac{1}{3}$ of their initial values.

SOME eight years ago were published[1] the results of experimental studies of the flow of gas-liquid mixtures through porous media. On the basis of these empirical results a hydrodynamic system of equations was developed[2] for describing and predicting the behavior of heterogeneous fluids when flowing through porous media. The applicability of these hydrodynamic equations to practical flow problems was illustrated by the analysis of a few simple cases of steady state flow and the numerical study of a linear system under transient flow conditions. The experimental work has since been extended to other porous media,[3] including several consolidated sands,[4] and to other multiphase fluid systems,[5] including[6] the simultaneous flow of oil, gas, and water through unconsolidated sands. Rather little, however, has been done in solving the basic differential equations and applying them to practical oil and gas producing systems. Additional applications[7] have been made to show how the productive capacity of oil wells can be calculated, and the effect on the productive capacity of the gas-oil ratio of the production and the connate water in the oil bearing formation. The basic principles of the well spacing problem were developed in an idealized form,[8] using the previously reported numerical studies of the transient flow in linear systems. And an application[9] has been made, too, in an analysis of the detailed mechanism of fluid

[1] R. D. Wyckoff and H. G. Botset, Physics **7**, 325–45 (1936).
[2] M. Muskat and M. W. Meres, Physics **7**, 346–63 (1936).
[3] G. L. Hassler, R. R. Rice, and E. H. Leeman, Trans. A. I. M. E. **118**, 116–37 (1936); and L. S. Reid, R. L. Huntington, Trans. A. I. M. E. **127**, 226–39 (1938).
[4] H. G. Botset, Trans. A. I. M. E. **136**, 91–103 (1940); H. Krutter and R. J. Day, Petr. Dev. and Tech. **6** (Nov. 1943).
[5] E. N. Dunlap, Trans. A. I. M. E. **127**, 215–25 (1938); M. C. Leverett, Trans. A. I. M. E. **132**, 149–69 (1939).
[6] M. C. Leverett and W. B. Lewis, Trans. A. I. M. E. **142**, 107–15 (1941).
[7] H. H. Evinger, and M. Muskat, Trans. A. I. M. E. **146**, 126–38, 194–203 (1942).
[8] M. Muskat, Trans. A. I. M. E. **136**, 37–52 (1940).
[9] S. E. Buckley and M. C. Leverett, Trans. A. I. M. E. **146**, 107–16 (1942).

displacement in the porous medium. However, the broad problem of general reservoir performance has only recently been attacked[10] from a quantitative point of view. On the other hand, these latter studies have involved trial and error procedures which are quite laborious. It is the purpose of this paper to treat this problem more directly.

The major significant result of the original fundamental experiments on the flow of gas-liquid mixtures through porous media was that such flow could be described by associating with each fluid phase a separate permeability. That is, the "law of force" for the composite multiphase system could be represented in terms of individual Darcy equations with permeability coefficients pertaining to the corresponding fluid phase. Thus considering the porous medium to contain all three of the fluid phases of general practical interest, namely oil, gas and water, their flow could be described by the set of equations:*

$$\bar{v}_o = -\frac{k_o}{\mu_o}\nabla p, \quad \bar{v}_g = -\frac{k_g}{\mu_g}\nabla p, \quad \bar{v}_w = -\frac{k_w}{\mu_w}\nabla p, \quad (1)$$

where $\bar{v}_o$, $\bar{v}_g$, $\bar{v}_w$ represent the vector flux velocities —flow rate per unit area—for the oil, free gas, and water, respectively, ∇p denotes the gradient in fluid pressure, μ_o, μ_g, μ_w are the viscosities of the individual phases, and k_o, k_g, k_w, are the permeabilities to the individual phases.

The point of importance about Eq. (1) is that instead of attempting to consider the multiphase system as a single complex fluid having a fictitious average density, viscosity, and equation of state, each phase can be treated separately. The interaction between the phases is expressed by the fact that the permeability coefficients k_o, k_g, and k_w each depend upon the volumetric phase distribution in the system. These latter may be expressed in terms of the fluid saturations of the individual phases, and will be denoted by ρ_o, ρ_g, ρ_w, expressed as fractions of the pore space. While the problem is thus reduced again to the Darcy law representation, it requires at the same time, however, that the permeability concept be broadly generalized. Whereas for homogeneous fluids the permeability played the role of providing a complete description of the structure of the porous medium with respect to its flow resistance, and otherwise was independent of the type of fluid, here, the permeability is seen to depend explicitly on the nature of the fluids and their distribution. The major empirical problem, in fact, thus lies in the determination of the relationship between the permeability to the individual phases and the fluid saturations. This relationship can be represented graphically by curves of the permeability to the individual phases as functions of the fluid saturations.

To illustrate the nature of these relationships it will suffice to note the results obtained[1] on gas-liquid mixtures, as shown in Fig. 1. Here the ordinates represent the permeability to the individual phases, expressed as percentages of the homogeneous fluid permeability. The abscissas denote the concentration of the liquid phase, in percent of the pore space. The rapid fall in liquid permeability as the liquid saturation decreases, and the corresponding rise in gas permeability, are the predominant features of these curves, and apply qualitatively to all two phase flow systems thus far studied. It is the values of k, either in analytical form or graphically, as given by these curves, and the corresponding ones for the water phase if the latter is of importance, which must be used in Eq. (1) to describe the heterogeneous fluid flow system. To complete the formulation of the hydrodynamic structure of such systems, Eq. (1) must be combined with the equations of continuity of the separate phases, and their equations of

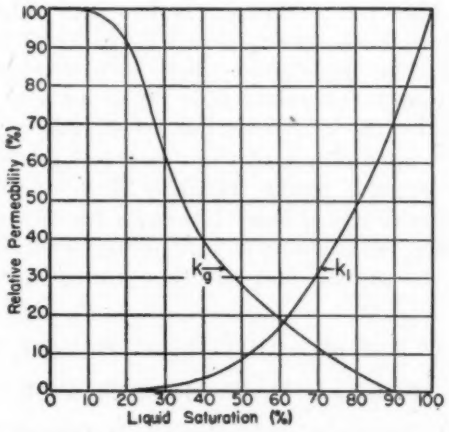


FIG. 1. Relative permeabilities for the liquid phase (k_l) and for the gas phase (k_g) of a gas-liquid mixture as a function of the liquid saturation. Relative permeability (percent) = (100 × actual permeability)/(homogeneous fluid permeability).

[10] E. C. Babson, Petr. Dev. and Tech. 7 (Jan. 1944); J. Tarner, Oil Weekly (June 12, 1944).

* For simplicity, the effects of body forces, as gravity, are not included here. Nor are the capillary pressures associated with the curvatures of the fluid interfaces, which in turn may be related to the fluid saturations.

state. In this way one obtains the equations:[11]

$$\nabla\cdot\left[\left\{\frac{S_o k_o}{\beta_o\mu_o}+\frac{S_w k_w}{\beta_w\mu_w}+\frac{\gamma k_g}{\mu_g}\right\}\nabla p\right] = f\frac{\partial}{\partial t}\left\{\frac{S_o\rho_o}{\beta_o}+\frac{S_w\rho_w}{\beta_w}+\gamma\rho_g\right\},$$

$$\nabla\cdot\left(\frac{k_o}{\beta_o\mu_o}\nabla p\right)=f\frac{\partial}{\partial t}(\rho_o/\beta_o), \qquad (2)$$

$$\nabla\cdot\left(\frac{k_w}{\beta_w\mu_w}\nabla p\right)=f\frac{\partial}{\partial t}(\rho_w/\beta_w),$$

$$\rho_o+\rho_g+\rho_w=1,$$

where S denotes the gas solubility, β the formation volume, that is, the volume of the liquid phase in the reservoir per unit volume of the same phase at atmospheric standard conditions, γ the free gas phase density, f the porosity, ρ the fluid phase saturation expressed as a fraction of the pore space, t the time variable, and where the subscripts o, g, and w again refer to the oil, gas, and water phases. The quantities k, μ, and p have the same meaning as in Eq. (1).

By the solution of these simultaneous partial differential equations the complete history of any multi-phase flow system can be described. As previously indicated, however, only a very small number of special cases have been thus far treated. Even a cursory inspection of these equations will readily show that they are of such high order of complexity that general solutions will be too difficult to derive. It is nevertheless possible to apply the basic physical principles underlying these equations to the study of gas drive oil producing reservoirs, that is, those in which the energy for oil expulsion is derived largely from the gas originally dissolved in the oil or that added to the formation in so-called "gas injection operations." A treatment of this type will be presented in the following sections.

THE DEPLETION HISTORY OF GAS-DRIVE RESERVOIRS

In this section will be considered the problem of determining the behavior of a gas-drive reservoir as it produces gas and oil without additional gas injection. In lieu of attempting to solve Eqs. (2) explicitly, the following simplified representation will be used. The reservoir will be considered as a uniform homogeneous porous medium containing oil, gas, and water in continuous equilibrium. The pressure in it will be taken as uniform, as if the reservoir were equivalent to a tank provided with continuously distributed outlets for fluid withdrawal. Under these conditions, the reservoir pressure can, in principle, be determined from a knowledge of the characteristics of the oil and gas, as represented by the quantities S, β, and γ, for any given conditions of total oil and gas withdrawal. These involve the so-called "material balance equations," which simply express the requirements of conservation of the total oil and gas contents, and constancy of the reservoir volume. Applying these conditions in differential form, one obtains a relation between the reservoir pressure, the oil and gas saturations in the sand, changes in the latter, and the gas-oil ratio. To proceed further, an appeal is then made to the fundamental permeability-saturation relationship, previously discussed, by imposing the requirement that the gas-oil ratio at any stage of the production must be compatible with the instantaneous reservoir fluid distribution, as implied by the permeability-saturation data for the rock. In this way, the gas-oil ratio is again related to the reservoir pressure and the fluid saturations. By combining these two relations, one finally obtains a differential equation for the variation of the reservoir oil saturation with the reservoir pressure, which, on integration, will give the pressure *vs.* cumulative oil recovery histories. At the same time, the gas-oil ratio *vs.* the pressure or cumulative recovery is readily obtained by applying again the permeability-saturation data. These, of course, represent the pertinent phases of the production history of the system in question.

While Eqs. (2) will not enter the analysis explicitly, they will not be directly violated. For Eqs. (2) are themselves nothing more than expressions of the "material balance," i.e. conservation requirements, in differential form, under conditions where there may be a fluid interchange between neighboring differential elements because of the pressure gradients in the system. Here these gradients will be neglected, so that each volume element can be treated independently and required to satisfy conservation conditions as if it were itself a closed reservoir. Accordingly, the only additional information which would be provided by the solution of Eqs. (2) would be the time behavior of the

[11] M. Muskat, Proc. Sec. Hydraul. Conf., Univ. of Iowa (1943).

system, and the *variation* in the local production histories of different elements in the reservoir, arising from the fact that the reservoir must actually be produced from individual wells rather than by a uniform and continuous distribution of fluid outlets. The present treatment makes no commitment on the former. And as to the latter, independent considerations show that the presence of wells as discrete fluid outlets should not materially influence the *average* behavior of the reservoir, except for localized phenomena about the well bores.

To formulate these physical statements* analytically, it is convenient to introduce the following additional notation. Q_o = rate of oil production, measured in stock tank barrels per day. Q_g = rate of gas production, in ft.^{3}/day, at standard conditions. $R = Q_g/Q_o$ = gas-oil ratio: ft.^{3}/bbl.

$$\alpha(p) = \frac{\gamma\beta\mu_o}{\mu_g}; \quad \Psi(\rho_o) = k_g/k_o.$$

Now it may be readily shown[2] that R will be given by:

$$R = \frac{Q_g}{Q_o} = S + \alpha(p)\Psi(\rho_o), \tag{3}$$

provided that the connate water which may be present is immobile. This equation gives expression to the requirement that the gas-oil ratio be compatible with the fluid saturations, as implied by the permeability-saturation relationship.

The oil content** of the reservoir per unit of volume of pore space will be, at any time:

$$\text{Oil content} = \rho_o/\beta. \tag{4}$$

Likewise, the gas content per unit volume of pore space will be:

$$\text{Gas content} = S\rho_o/\beta + \gamma(1 - \rho_w - \rho_o). \tag{5}$$

Assuming now that the solubility of gas in the connate water can be neglected, the oil and gas production will be given by:

$$Q_o = \Delta(\rho_o/\beta), \tag{6}$$

$$Q_g = \Delta[S\rho_o/\beta + \gamma(1 - \rho_w - \rho_o)], \tag{7}$$

where Δ indicates an increment in a unit time interval. Eqs. (6) and (7) represent the material balance or conservation requirements in differential form.

They also lead to an expression for the gas-oil ratio, namely:

$$R = S + \frac{\frac{\rho_o}{\beta}\Delta S + (1 - \rho_w - \rho_o)\Delta\gamma - \gamma\Delta\rho_o}{\Delta(\rho_o/\beta)}. \tag{8}$$

Equating this to Eq. (3), expressing the differentials in terms of pressure variation, and introducing the notation:

$$\lambda(p) = \frac{1}{\gamma\beta}\frac{dS}{dp}; \quad \epsilon(p) = -\frac{1}{\gamma}\frac{d\gamma}{dp}; \quad \eta(p) = \frac{\alpha}{\beta^2\gamma}\frac{d\beta}{dp}, \tag{9}$$

$$\omega(\rho_o, p) = 1 + \frac{\mu_o}{\mu_g}\Psi(\rho_o),$$

one obtains:

$$\omega(\rho_o, p)\frac{d\rho_o}{dp} = \rho_o\lambda(p) + (1 - \rho_w - \rho_o)\epsilon(p) + \rho_o\Psi(\rho_o)\eta(p). \tag{10}$$

This is the final differential equation for the production history of the system. By integrating it numerically, one obtains directly the relation between the residual oil saturation ρ_o and the reservoir pressure p. On inserting the values so obtained into Eq. (3), the corresponding gas-oil ratios are readily computed. The cumulative recovery in terms of stock tank oil corresponding to the residual oil saturations are then calculated from the relation:

$$\text{Cumulative recovery} = \left(\frac{\rho_o}{\beta}\right)_i - \frac{\rho_o}{\beta}, \tag{11}$$

referred to unit volume of pore space, the subscript i denoting the initial values of ρ_o and β.

To provide a basis of reference in the discussion of gas injection systems a numerical application of Eqs. (10) and (11) will first be made to a hypothetical reservoir originally at 2500 lb. pressure and producing oil entirely by solution

* As mentioned earlier, Babson and Tarner (Reference 10) used essentially the same physical representation. However, instead of expressing the "material balance" conditions in differential form and constructing a differential equation, as will be done here, they expressed the conservation requirements in integrated form, i.e. in terms of cumulative oil and gas production, and were forced to resort to a trial and error numerical treatment. Moreover, Tarner, who did explicitly treat the problem of gas injection, considered only the case of no gravity segregation. In addition to the latter, also the case where there is complete gas segregation will be treated here.

** In expressing conservation requirements, the fluid volumes must be expressed in terms of their stock tank, or standard conditions, equivalents.

gas drive. The characteristics of the oil and gas will be assumed to be those shown in Fig. 2. Here are plotted curves of variation of gas solubility, the expansion factors* for the oil, the viscosity of the oil and gas, and gas density, all as functions of the reservoir pressure. It will be noted that the gas solubility corresponds to a solution gas-oil ratio of 535 ft.³/bbl. These data permit a direct evaluation of the functions $\lambda(p)$, $\epsilon(p)$, $\eta(p)$ and the ratio μ_o/μ_g entering into Eq. (10).

The permeability saturation relationship, as defined by the function $\Psi(\rho_o)$, was taken to be that shown in Fig. 3, for the case where the sand has a connate water saturation of 30 percent. This curve was obtained from the data published by Leverett and Lewis.[6] The relative permeability to the oil—the actual permeability divided by the homogeneous fluid permeability—is also plotted in this figure. With these data at hand, the numerical integration of Eq. (10) is quite straightforward and simple to carry out. The results are shown in Figs. 4 and 5. In Fig. 4 are plotted the variations of residual oil saturation in the reservoir and the gas-oil ratio of the production as a function of the reservoir pressure. In Fig. 5 the results are plotted as functions of the cumulative oil recovery, as calculated by Eq. (11). In this figure, in addition to the gas-oil ratio history, the reservoir pressure variations with cumulative recovery are also shown. In order to show explicitly the practical significance of the cumulative oil recoveries, an additional abscissa scale has been indicated in Fig. 5, translating the values of percentage of pore space into those of stock tank barrels per acre, on the assumption that the pay has a thickness of 40 feet and a porosity of 25 percent.

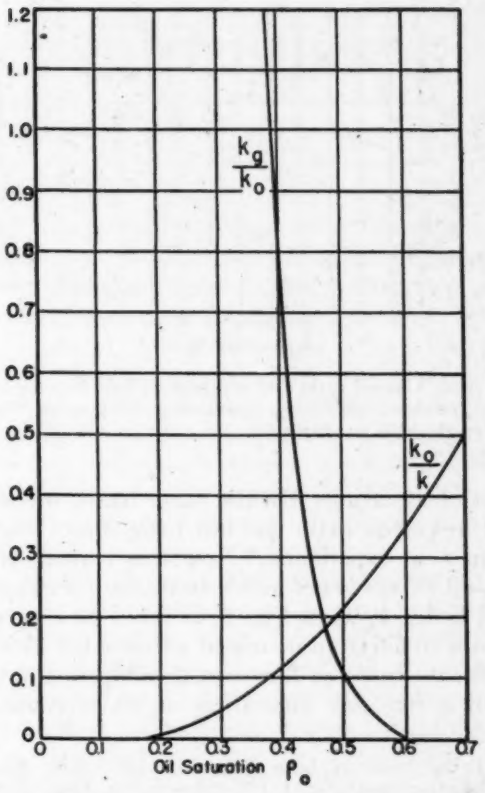


FIG. 3. Permeability saturation curves used in the calculation of gas-drive reservoir histories. $\Psi(\rho_o) = k_g/k_o$ = (permeability to gas)/(permeability to oil). Connate water = 30 percent; k_o/k = (permeability to oil)/(homogeneous fluid permeability).

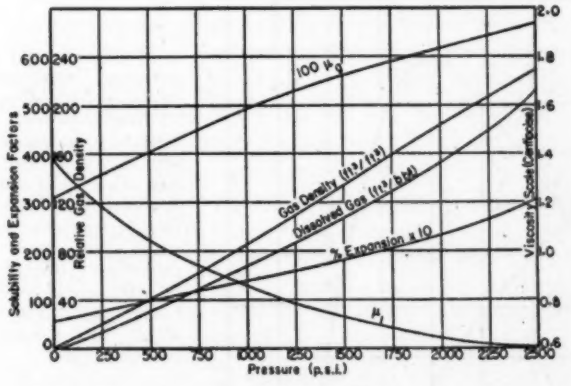


FIG. 2. Physical characteristics of oil and gas assumed in the calculation of gas-drive reservoir histories. μ_l, μ_g are viscosities at reservoir temperature of oil and gas; expansion factors are equivalent to $\beta - 1$.

* These are numerically equal to the formation volume factors, β, minus one, i.e. the excess of the reservoir volume over unit stock tank volume.

As these curves are quite similar to those previously computed by Babson and Tarner, they will be given no extended discussion here. It will suffice to point out that without making any explicit assumption regarding the total gas-drive recovery, the analysis automatically leads to a cumulative stock-tank oil recovery equivalent to 14.5 percent of the pore volume, and a reservoir space voidage of 28.7 percent. While the assumed basic reservoir data, as shown in Figs. 2 and 3, do not refer to any actual single producing field, but include typical observed fluid characteristics from several sources, it is of interest to note that the corresponding recovery predictions are of the same order of magnitude as commonly found in practice. Moreover, the rise in the computed gas-oil ratio history, and ultimate decline, as plotted in Figs. 4 and 5, also simulate

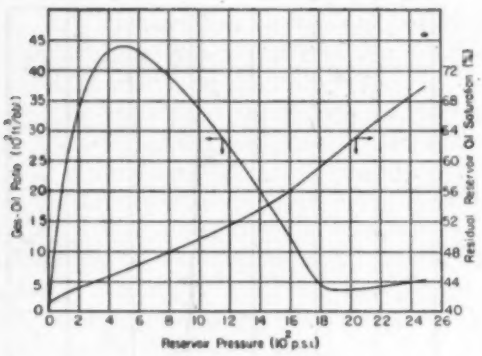


FIG. 4. Gas-oil ratio and residual oil histories for a solution gas-drive reservoir, containing 30 percent connate water, with no gas injection.

field observations. On the other hand, the initial fall in gas-oil ratio has but little direct support from field experience.* Now this feature of the gas-oil ratio curves arises from the fact that k_g, and hence k_g/k_o in Fig. 3, is taken as zero until ρ_o falls to 0.6 from its initial value of 0.7. Accordingly, no free gas flow can develop in the sand until a free gas saturation of 10 percent—the "equilibrium gas saturation"—has been built up, during which time the gas-oil ratios will be merely the solution ratios, which fall as the reservoir pressure declines. Failure to observe the initial fall in gas-oil ratio may, therefore, imply that in actual consolidated rocks comprising oil producing strata, the equilibrium gas saturation required for free gas flow is appreciably less than the 10 percent assumed here. Or it may mean that during the initial history local well conditions actually mask the general reservoir behavior.

* It has, however, been observed in laboratory reservoir model experiments (cf. I. I. Gardescu, Pet. Engr. *4*, Nov., Dec., Jan., and Feb., 1932–33).

GAS INJECTION WITH NO SEGREGATION

For this case, which is essentially that treated by Tarner, the gas cap overlying the oil zone, if originally present, is assumed to maintain a constant volume. The injected gas, plus that in the gas cap initially, is supposed to diffuse through and be produced with the oil zone, except for that required to maintain pressure equilibrium throughout the system. Following a procedure similar to that outlined for the case of straight depletion, it is found that the present type of gas injection operations will be governed by the differential equation:

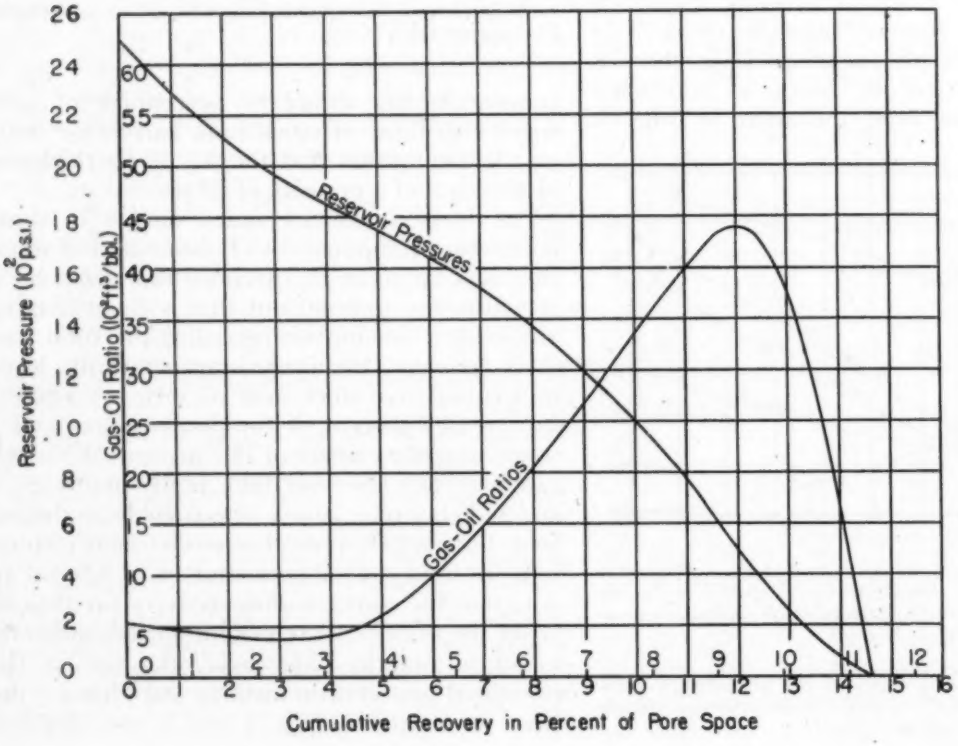


FIG. 5. Gas-oil ratio and reservoir pressure histories, as functions of cumulative oil recovery, for a solution gas-drive reservoir, without gas injection. Upper abscissa scale gives the cumulative recovery in 10^3 stock tank barrels/acre, for a 40 ft., 40 percent porosity, pay.

$$\frac{d\rho_o}{dp}=\frac{\rho_o\lambda(p)+\{\bar{H}(1-\rho_w-\beta\bar{\rho}_{oi})+1-\rho_w-\rho_o\}\epsilon(p)-\bar{H}\frac{\bar{\rho}_{oi}}{\gamma}\frac{d\beta}{dp}+\left(\Psi-\frac{r}{\alpha}R\right)\rho_o\eta(p)}{1+\frac{\mu_o}{\mu_g}\left(\Psi-\frac{r}{\alpha}R\right)} \tag{12}$$

where ρ_o, ρ_w, p, β, μ_o, μ_g, γ, α, Ψ, $\lambda(p)$, $\epsilon(p)$, $\eta(p)$ have the same meaning as before, r is the fraction of produced gas which is returned to the formation, R the producing gas-oil ratio, and $\bar{H}$ is the ratio of the thickness of the original gas cap zone to the oil-saturated zone. The connate water saturation, ρ_w, is assumed to be the same in the gas cap as in the oil zone. $\bar{\rho}_{oi}$ is the initial or residual oil saturation in the gas cap, expressed in stock tank equivalent, it being assumed that this oil is immobile and not further depleted by the operations, except for shrinkage.

The cumulative recovery per unit volume of original pore space of the saturated oil pay is given again by:

$$\text{Cumulative recovery}=(\rho_o/\beta)_i-\rho_o/\beta, \tag{13}$$

where the subscript i refers to the initial values.

Using the basic physical data of Figs. 2 and 3, Eq. (12) was integrated for several values of gas injection ratio r, and for the case $\bar{H}=0$, i.e., no original gas cap.* The gas-oil ratios were computed by Eq. (3) and cumulative oil recoveries by Eq. (13). In Fig. 6 are plotted the pressure and gas-oil ratio histories, *vs.* the cumulative stock tank recoveries for values of $r=1.0$, 0.8, 0.6, and 0, the curves for $r=0$ being those previously plotted in Fig. 5 for direct pressure depletion. The recoveries are expressed as fractions of the pore space.

These curves are similar to those plotted in Figs. 7 and 8 of the paper by Tarner. It will be seen that, as is to be expected physically, the pressure decline is increasingly retarded and the cumulative oil recovery is increased as more gas is returned to the formation. In the limiting case where all the produced gas is returned, the calculations for injection were terminated at the stage where the gas-oil ratio reached the value 20,000 ft.³/bbl., as further injection would very likely be unprofitable. It was assumed that after injection was stopped, the reservoir pressure was allowed to decline as a normal depletion process. Tarner indicates that complete recovery would be ultimately derived if all the gas is returned. This, however, will be true only if the oil permeability remains non-vanishing up to zero oil saturation. While such a situation is in itself extremely unlikely, the practical problem of operating under indefinitely increasing gas-oil ratios will anyway force a termination of the gas injection program before the ultimate limit is reached.

It will be observed that the ultimate oil recoveries, when expressed as percentages of the pore space, will be 14.5, 18.7, and 21.8 percent for $r=0$, 0.6 and 0.8. For $r=1$, the cumulative recovery will be 18.1 percent by the time the gas-oil ratio reaches 20,000 ft.³/bbl., and the final

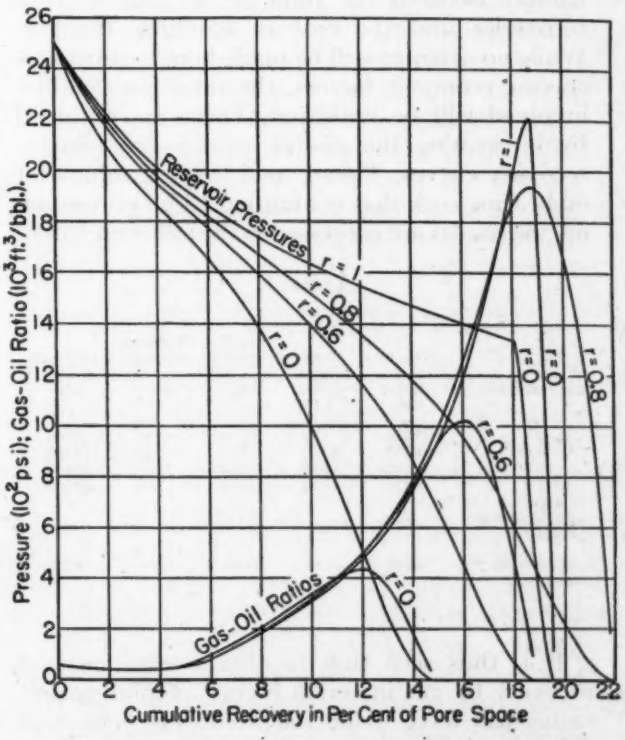


FIG. 6. Pressure and gas-oil ratio histories of gas-drive fields for various degrees of gas injection, with no gas segregation. r = fraction of produced gas returned to formation. For $r=1$, gas injection discontinued after 20,000 ft.³/bbl. gas-oil ratio is reached.

* While Eq. (12) presupposes that $\bar{H}$ retains its initial value, whatever it may be, the use of the term "no gas segregation" for this case refers to the assumption that the gas cap gas does not remain segregated, and that there is no gravity drainage causing a shrinkage of the oil zone, even though the initial gas cap will retain its identity as such.

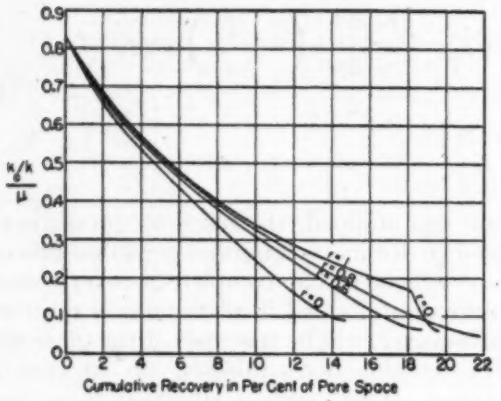


FIG. 7. The variation of the relative productivity factors —production rate per unit pressure differential—during the producing life of gas injection systems with no gas segregation. k_0/k = (permeability to oil)/(homogeneous fluid permeability); μ = oil viscosity under reservoir conditions; r = fraction of produced gas returned to formation.

recovery will be 19.6 percent after ultimate depletion of the reservoir pressure. These results show the effectiveness of gas injection from a physical point of view. The economic value of the operations, however, will depend on the balance between the value of the increased oil recoveries and the cost of handling the gas. While no attempt will be made here to introduce specific economic factors, the total gas volumes involved will be indicated. These are obtained by integrating the gas-oil ratio *vs.* cumulative recovery curves. Taking as a unit of volume of oil bearing rock that containing 1 bbl. of residual oil, the results are expressed in the following table:

TABLE I.

Fraction of gas returned, r	Vol. oil produced bbls.	Total gas produced ft.³/bbl.	Gas injected per bbl. recovery: ft.³	Effective average gas-oil ratio: ft.³/bbl.	Maximum gas-oil ratio ft.³/bbl.
0	0.271	534	0	1959	4400
0.6	0.350	1270	2177	3629	10300
0.8	0.407	1459	2868	3585	19450
1.0 (to a gas-oil ratio of 20,000 ft.³/bbl.)	0.337	2590	7754	7754	20000
0 to ultimate depletion.	0.366	2746	7067	7503	22150

It is thus seen that to obtain the increased recovery by gas injection higher average gas-oil ratios will have to be tolerated. The indicated difference in average gas-oil ratio between $r=0.8$ and 0.6 is probably within the uncertainties of the numerical computations and of no significance. On the other hand, it should be noted that the values of the last row, for complete gas return, appear so unfavorable only because it was assumed that the injection is discontinued at a gas-oil ratio of 20,000 ft.³/bbl. If a lower limit had been set for the maximum gas-oil ratio the recovery for $r=0.8$ would also have been drastically reduced. Moreover, in the case of $r=1$, the gas-oil ratio during the natural depletion period rises appreciably above 20,000 ft.³/bbl. before it begins to decline. In fact, it is because of the continued high gas-oil ratios after injection is stopped that the pressure drops so abruptly with only a small increase in oil recovery. Of course, if instead of completely discontinuing the gas injection at the limiting gas-oil ratio, the fraction returned were reduced so as to limit the total volume rate of injection,* a more gradual decline in pressure would result, and a correspondingly greater recovery.

The reason for the continued rise in gas-oil ratio, and apparently less efficient operation, for the higher values of r, is that the maintenance of higher reservoir pressures gives a greater atmospheric gas volume equivalent of the free gas flow through the formation. The ultimate decline in gas-oil ratio is accordingly retarded. However, if the operations are continued beyond the maximum gas-oil ratio, the increase in cumulative recovery becomes readily apparent.

Because of the gas evolution within the pay as the pressure declines, the permeability to the oil flow will decrease as production proceeds. At the same time the oil viscosity will increase. The result will be a continuous decline in productivity factor for the producing wells. As the productivity factor is proportional to the ratio of the permeability to the viscosity,** this ratio is plotted in Fig. 7 as a function of the cumulative oil recovery. The values of oil permeability were taken from Fig. 3. It will be seen that while the ultimate value of the productivity factor will decrease with increasing recovery, in all cases it will be of the order of $\frac{1}{10}$ of that at the beginning of production.

* For operations in which the gas is injected at a preassigned ratio, $\bar{r}$, with respect to the oil produced, one need only replace in Eq. (12) rR by $\bar{r}$.

** This refers to the limiting case of vanishing pressure differential, where the variation of k_0/μ through the pay can be neglected. The productivity factor itself is defined as the flow rate per unit pressure drop over the sand. When referred to unit sand thickness it is termed the "specific" productivity factor. Moreover, the values of $(h_0/h)/\mu$ plotted in Fig. 7, and Figs. 11 and 12 below, give a measure only of the rate of oil flow into the bore hole, and should be divided by β to get rates at the surface.

GAS INJECTION WITH COMPLETE GRAVITY SEGREGATION

By the nature of the basic assumption underlying this type of operation, the mechanics of the production is essentially one of gravity drainage. The injected gas serves merely to maintain the reservoir pressure. Indeed, this type of operation represents the idealized form of "pressure maintenance." The value of such pressure maintenance is in sustaining higher production rates and in prolonging the flowing life of the wells. The ultimate oil recovery, however, will be essentially independent of the fractional return of the gas. It is determined by the value of the residual oil left by gravity drainage, it being assumed that the rate of production is sufficiently restricted to permit the segregation between the gas and oil to keep pace continuously with the oil withdrawal from the oil producing section. It is, of course, also presupposed that as the gas cap expands and the gas-oil interface is lowered the producing interval in the well bore is continually reduced to lie below the gas-oil contact. Accordingly, the injected gas remains trapped in the gas cap.

The operations are completed when the gas-oil contact reaches the bottom of the original oil-saturated section. The residual oil saturation in the reservoir is simply that characteristic of the gravity drainage process, plus that which may have been present in the initial gas cap, and which is assumed to be immobile throughout the production history. The ultimate cumulative recovery is, therefore, merely the difference between the residual oil and the initial stock tank oil in the oil producing section. Of course, this implies the validity of the various assumptions already indicated. It is admittedly doubtful if conditions justifying all these assumptions will often occur in practice. However, as at least the *tendency* for gravity segregation will always be present in producing fields, it is of interest to determine the nature of the production histories that would obtain if the effect of gravity drainage could be given full play.

To crystallize the nature of the production mechanism under consideration, the reservoir may be visualized as a simple tank, as diagrammatically indicated in Fig. 8. The gas cap will occupy the upper part of the section to a fractional thickness $\bar{h}$. Its initial thickness will be denoted by $\bar{h}_i$. The further new notation will be:

$\bar{\rho}_{og}$ = residual (stock tank) oil saturation in the drained gas cap, i.e., in $(\bar{h}-\bar{h}_i)$.

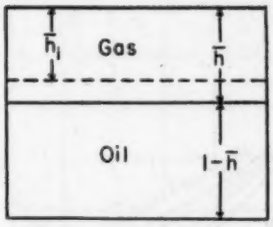


FIG. 8. Diagrammatic representation of a solution gas-drive reservoir containing a gas cap, of fractional thickness $\bar{h}$, with complete gas segregation.

$\bar{\rho}_{oi}$ = residual (stock tank) oil saturation in the initial free gas cap, i.e., in $\bar{h}_i$; $\zeta=\gamma\beta-S$.

By applying the requirement of gas conservation to the system, it is readily found that the gas cap thickness $\bar{h}$ will be given by:

$$\bar{h}\left[\gamma(1-r)(1-\rho_w)+\zeta\left(\frac{r\rho_o}{\beta}-\bar{\rho}_{og}\right)\right]$$
$$=\gamma_i(1-\rho_w)\{r(1-\bar{h}_i)+\bar{h}_i\}-\bar{\rho}_{oi}\zeta_i\bar{h}_i$$
$$-r(1-\bar{h}_i)\rho_{oi}\zeta_i/\beta_i-(\bar{\rho}_{og}-\bar{\rho}_{oi})\bar{h}_i\zeta$$
$$-r\gamma(1-\rho_w)+r\rho_o\zeta/\beta, \quad (14)$$

where the symbols γ, r, ρ_w, ρ_o, β have the same meaning as before, and the subscript i indicates the initial value. In addition to the assumptions already discussed, it has been further assumed, in deriving Eq. (14), that the connate water is immobile and the saturation ρ_w has the same constant value throughout both the gas cap and oil zone. While strict uniformity of the connate water saturation is very unlikely in actual producing formations, it is virtually necessary to assume such uniformity in lieu of specific data to the contrary. An equally arbitrary assumption is the constancy of the residual oil saturation in the drained gas cap section, i.e., $\bar{\rho}_{og}$. It would probably be more accurate to take the residual formation or reservoir volume, immediately above the gas-oil contact, i.e., $\beta\bar{\rho}_{og}$, as constant. However, this would materially complicate the analysis with but little effect on the final results.

Equation (14) in itself evidently cannot provide a complete description of the production history. For it expresses only a relation between $\bar{h}$ and the variables ρ_o and p, the latter through the variables γ, ζ, and β. The additional link between these variables is given by the observation that since the oil zone itself continues to produce by the gas-drive mechanism, the producing gas-oil ratio will be governed by Eq. (3), and the relation between ρ_o and p by Eq. (10), as in the case of the simple gas drive reservoir. Accordingly, Eq. (14) actually will serve to relate the gas cap

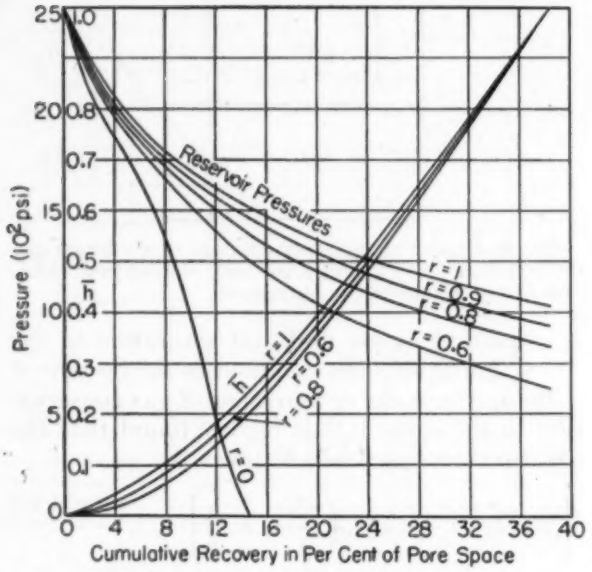


FIG. 9. Histories of reservoir pressure and position of gas-oil contact in gas-drive fields for various degrees of gas injection, assuming complete gas segregation. r = fraction of produced gas returned to the formation; $\bar{h}$ = fraction of formation occupied by gas cap.

thickness $\bar{h}$ to either the reservoir pressure p, or the oil saturation in the oil producing zone ρ_o. Moreover, this relationship may be expressed in terms of the cumulative oil recovery, on noting that the latter will be given by:

$$\text{Cumulative recovery} = (1-\bar{h}_i)(\rho_o/\beta)_i - (1-\bar{h})\rho_o/\beta - (\bar{h}-\bar{h}_i)\bar{\rho}_{og}. \quad (15)$$

The calculations are terminated at the values of ρ_o or p at which $\bar{h}$ becomes equal to unity, which implies that the gas cap has expanded throughout the whole of the oil zone. The ultimate recovery will then be:

$$\text{Ultimate recovery} = (1-\bar{h}_i)\{(\rho_o/\beta)_i - \bar{\rho}_{og}\}. \quad (16)$$

As previously indicated, this will be independent of the fractional gas return, except when the latter is so small that the reservoir pressure falls to atmospheric before the gas-oil contact has been depressed to the bottom of the oil zone.* This, however, will occur only when the amount of gas returned is such a negligible fraction of that produced that the operations would hardly be considered as pressure maintenance from a practical point of view. In fact, the final reservoir pressure reached when $\bar{h}=1$, or that injection ratio r which will result in a given termination pressure, may be computed from the equation:

$$r(1-\bar{h}_i)\left\{\gamma_i(1-\rho_w)-\frac{\rho_{oi}\zeta_i}{\beta_i}\right\} = h_i\bar{\rho}_{oi}\zeta_i - \{\bar{h}_i\bar{\rho}_{oi}+(1-h_i)\bar{\rho}_{og}\}\zeta + (1-\rho_w)\{\gamma-\gamma_i\bar{h}_i\}. \quad (17)$$

For the special case where the gas injection begins in a virgin pay without any initial gas cap, Eq. (14) reduces to:

$$\bar{h}=\frac{r[(\gamma_i-\gamma)(1-\rho_w)+\rho_o\zeta/\beta-(\rho_o\zeta/\beta)_i]}{\gamma(1-r)(1-\rho_w)+\left(\frac{r\rho_o}{\beta}-\bar{\rho}_{og}\right)\zeta}. \quad (18)$$

For this case, calculations of reservoir pressure and the position of the gas-oil contact were made for several values of the gas injection ratio r. The basic reservoir data were the same as used in the previous discussion, and as shown in Figs. 2 and 3. In addition, the value used for $\bar{\rho}_{og}$, i.e., the residual (stock tank) oil saturation after gravity drainage, was 0.15.** From Eq. (16) this immediately fixes the ultimate recovery as 0.3852, in terms of the pore space, for values of r

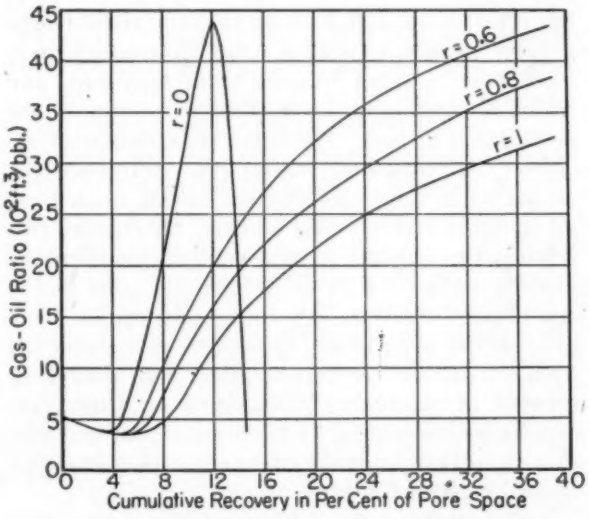


FIG. 10. Gas-oil ratio histories of gas-drive reservoirs for various degrees of gas injection, assuming complete gas segregation. r = fraction of produced gas returned to the formation.

* If the production by gravity drainage alone can be continued, the ultimate recovery will even then equal the value given by Eq. (16).

** This value, which is of the same order as that assumed for the water flooding mechanism, may seem to be too low. However, such values have been obtained in unconsolidated sand experiments. Moreover, it is used here only for illustrative purposes, rather than as a well-established empirical result.

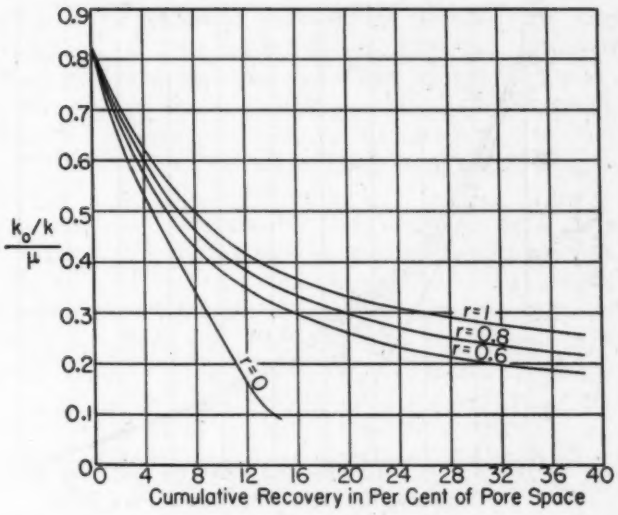


FIG. 11. The variation of the relative specific (per unit thickness) productivity factors during the producing life of gas injection systems, with complete gas segregation. k_o/k = (permeability to oil)/(homogeneous fluid value); μ = oil viscosity under reservoir conditions; r = fraction of produced gas returned to formation.

exceeding 0.011. The results are plotted in Fig. 9.

The implications of Fig. 9 are quite obvious. As previously discussed, the ultimate oil recovery is the same for the various values of r, except the limiting case of no gas return, i.e., $r=0$. The immediate effect of gas injection lies only in the retardation of the pressure decline, thus prolonging the flowing lives of the producing wells, and permitting greater production rates. It will be observed, also, that the lowering of the gas-oil contact, i.e., variation of $\bar{h}$, is determined largely by the cumulative recovery, and is affected but little by the value of r.

The gas-oil ratio history of the oil section itself will be the same as for the simple solution gas drive case ($r=0$), when expressed as a relationship between the gas-oil ratio and reservoir pressure. However, because of the different rates of pressure decline, when plotted against the cumulative oil recovery they give a relative behavior as shown in Fig. 10. These gas-oil ratios reflect only the inherent producing characteristics of the oil saturated section. Superposed on the depletion of oil and gas saturation in this section will be the contributions in gas and oil production caused by the shrinkage in thickness of the oil producing stratum as the gas cap expands and the gas-oil contact is lowered. These contributions must be added to those included in Fig. 10 to give the resultant ratios of the system as a whole.

The fact that the curves for $r>0$ in Fig. 10 do not show the maximum and fall, as does that for $r=0$, is evidently the effect of the maintenance of pressure. Thus referring to Fig. 9, the pressure will fall to only 1035 p.s.i., for $r=1$, even by the time all the recoverable oil has been produced. For $r=0.9$ the final pressure is 937 p.s.i., for $r=0.8$ it is 837, and for $r=0.6$ it is 630 p.s.i. By reference to Fig. 4 it will be seen that the gas-oil ratios for the producing section will not reach a maximum until the reservoir pressure has fallen to 500 p.s.i. Hence, as long as the fraction of gas returned exceeds that giving an ultimate reservoir pressure of 500 p.s.i. the gas-oil ratio will continue to rise until the operations are completed.

Because of the oil depletion in the oil zone and increasing oil viscosities the productivity factors will here, too, decrease during the productive life. In addition, there will be superposed here the effect of shrinkage of the oil zone as the gas cap expands. In Fig. 11 are plotted the values of $(k_o/k)/\mu$, which give the relative *specific* productivity factors *vs.* the cumulative oil recovery. In Fig. 12 are plotted the relative values of the resultant productivity factors, taking into account the varying thickness of the oil zone according to the curves of Fig. 9.

It will be noted, by comparing Fig. 11 and Fig. 7, that the specific productivity factors do not fall quite as much when there is gas segregation as when there is no gravity drainage. Moreover, the decrease in productivity factor will be greater for the smaller values of r. On the other hand, when the decreasing oil zone thickness is taken into account, Fig. 12 shows that the resultant productivity factors will finally all fall to zero. This effect, of course, will ultimately counterbalance the tendency for sustained higher production rates because of the retardation in

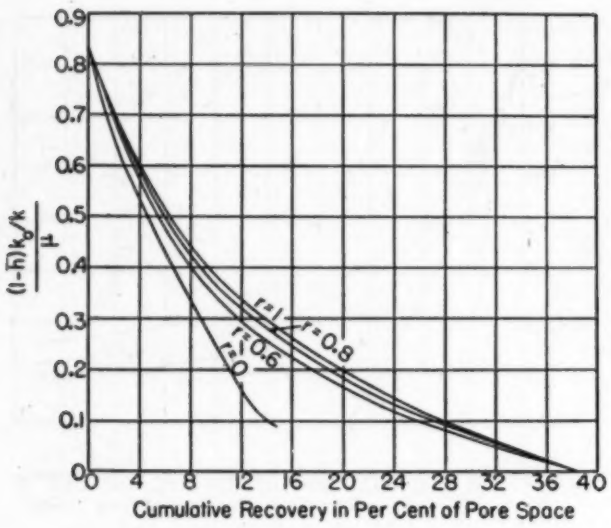


FIG. 12. The variation of the productivity factor during the producing life of gas injection systems with complete gas segregation. k_o/k = (permeability to oil)/(homogeneous fluid permeability); μ = oil viscosity under reservoir conditions; $\bar{h}$ = fractional thickness of gas cap; r = fraction of produced gas returned to the formation.

reservoir pressure decline due to the gas injection.

CONCLUSIONS

The theoretical analysis of the performance of gas-drive oil producing reservoirs leads to predictions of pressure and gas-oil ratio histories and ultimate oil recoveries which, at least qualitatively, are in agreement with those observed in practice. Under direct pressure depletion—no gas injection—the gas-oil ratios initially decline, then rise rapidly to a maximum, and ultimately fall sharply on approaching the end of the production history. The reservoir pressure declines continuously, but at rates—in terms of cumulative recovery—which decrease as the gas-oil ratio decreases. In fact, the variation in pressure drop per unit volume of oil recovery shows a close parallelism to the gas-oil ratio curve. The physical ultimate oil recovery is found to be 14.5 percent of the pore space for the type of reservoir conditions assumed. This will result in an ultimate space voidage of 28.7 percent, and be equivalent to a recovery of 27.1 percent of the initial stock tank oil content of the reservoir.

The effect of gas injection on gas-drive reservoirs will depend on the distribution of the injected gas. The extreme forms of such distributions are: (1) a continuous diffusion through and production with the oil of the oil saturated zone, and (2) a complete segregation and trapping in the gas cap overlying the oil saturated zone.

In the first case, oil production takes place by a gas drive mechanism similar to that of direct depletion types of reservoir. The injected gas serves to retard the pressure decline and maintain the reservoir oil in an inflated state so as to minimize the gas permeability. By virtue of the continued gas circulation it will also increase the flushing of the oil from the pores of the pay. The resultant increased oil recovery will be accompanied by higher gas-oil ratios. If the operations can be continued through the period where the gas-oil ratios reach their maximum values, large increases in ultimate recovery can be obtained. For the hypothetical gas-drive reservoir analyzed in detail, it is found that as compared to the recovery of 27.1 percent of the original stock tank oil in place without gas injection, the recovery will be increased to 35.0 percent if 60 percent of the produced gas is returned, and to 40.7 percent if 80 percent is returned. In the latter case, however, the injection must be continued through a stage where the gas-oil ratios rise to 19,450 ft.^{3}/bbl. If all the gas is returned, the gas-oil ratios tend to rise indefinitely as production continues. By the time a ratio of 20,000 ft.^{3}/bbl. is reached, 33.7 percent of the original stock tank oil will be recovered. If injection is stopped then, and the pressure is allowed to decline, an additional yield of 2.9 percent will be obtained.

If complete gas segregation is assumed, the oil recovery will be essentially determined by the value of the residual oil left in the pay after gravity drainage. Assuming this is equivalent to a stock tank residual of 15 percent of the pore space, the ultimate recovery for the reservoir previously considered will be 72 percent of the original stock tank oil in the formation. The effect of the gas injection is largely that of maintaining the reservoir pressure while the gravity

drainage is taking place. Thus, if all the gas is returned, the reservoir pressure will be 1035 p.s.i. when the gas-oil contact will reach the bottom of the oil zone. If 60 percent of the gas is returned the corresponding final pressure will be 630 p.s.i. The rate of expansion of the gas cap and depression of the gas-oil contact is determined largely by the cumulative oil recovery rather than by the rate of gas return. The gas-oil ratios will rise due to the fall in saturation in the oil zone, though they will be limited in magnitude to those characterizing the direct depletion process.

In both types of gas injection systems the permeability to the oil will decrease and oil viscosity will increase as the oil saturation is lowered and the pressure declines. Moreover, the same types of change in permeability and oil viscosity will occur when no gas is returned to the formation. In all cases the result will be a continued fall in the productivity factors of the producing wells throughout the life of the field. When there is no gas segregation, the productivity factor at complete depletion will decrease with increasing ultimate recovery, finally falling to values 5–10 percent of the initial productivity factor. If the injected gas is trapped in the gas cap, the maintenance of the reservoir pressure will lessen the increase in viscosity and fall in the specific productivity factors to 20–30 percent of their initial values. However, when the decreasing thickness of the oil zone—caused by the gas cap expansion—is taken into account, the resultant productivity factors will in all cases fall to zero when the gas-oil contact reaches the bottom of the oil pay.

Unfortunately it is not possible at present to specify the precise conditions under which either of the two types of production mechanism discussed here will obtain in practice. Undoubtedly each will play a role in the great majority of actual gas injection operations. It seems likely that the case of complete gas segregation will in general require such low production rates, or pressure differentials over the pay, that it will seldom be practicable to operate a field in this manner. However, the results presented here on the recoveries which such production mechanisms can yield should provide a better basis for evaluating and planning the operating procedures.

The author wishes to thank Dr. A. G. Loomis of the Shell Development Company for the opportunity of discussing with him some of the phases of the general problem treated here, and Dr. Paul D. Foote, Executive Vice President of the Gulf Research & Development Company, for permission to publish this paper.

The Effects of Temperature and Humidity on the Physical Properties of Tire Cords

J. H. Dillon and I. B. Prettyman
The Firestone Tire and Rubber Company, Akron, Ohio

(Received October 31, 1944)

Apparatus by means of which tire cords may be conditioned and tested at elevated humidities (1.6–65 percent R. H.) and temperatures (20–165°C) is described. The apparatus consists of a portable conditioning unit which is used with a standard tensile testing machine. Equipment for measuring creep of tire cords under dead load at elevated temperatures is also described. Five representative types of tire cord were employed in the experiments, medium stretch cotton, low stretch cotton, viscose rayon, Fortisan, and Nylon. Tenacity, "10 pound-stretch," and ultimate stretch are studied as functions of relative humidity, temperature, and moisture regain. It is shown that the tensile properties are best represented as functions of temperature at constant regain. The tenacity of viscose rayon cords decreases with increasing temperature at constant regain in the range of regains 0–5 percent. The tenacity of the medium stretch cotton cords at constant regain falls off with increasing temperature in the range of 20–100°C and then changes very little up to 165°C. That of the low stretch cotton cords, however, decreases in a manner similar to that found for rayon. The Fortisan cords show a decrease of tenacity in the range 20–100°C and then a slight increase for higher temperatures. The cotton, rayon, and Fortisan cords give nearly linear creep curves (elongation *vs.* log time) in the range of creep times 0.002 to 20 hours. The creep curves for Nylon show a tendency to increase in slope at extended creep times. The creep data are analyzed in terms of two arbitrarily defined indices, "initial compliance" and "weighted creep." Both the tensile and creep characteristics are discussed from the viewpoint of current theories of structure and also in the light of their relations to serviceability in tires.

THE modern pneumatic tire is commonly classified as a rubber product. Its rubber tread satisfies the requirements of resistance to abrasion, cracks, cuts, and bruises while the

rubber ply and bead stocks serve to bind its structure together into a tough, resilient unit. The rubber inner tube serves as a resilient container for the inflating air. A pneumatic tire, however, would be a flabby, useless thing without the strong steel bead wire which holds it to the wheel rim and the cord fabric which supports the major portion of the inflation stresses, but, unlike the bead wire, must also withstand repeated flexure through large amplitudes. The

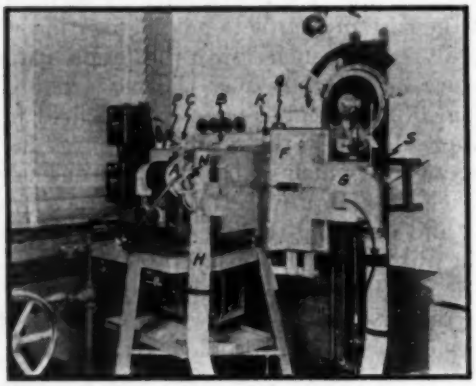


FIG. 1. Cord humidifier—general view.

cord fabric, in fact, must serve in the capacity of a flexible, resilient, and yet strong skeleton; resistant to heat, mechanical fatigue, and chemical attack and with a minimum of creep under sustained inflation stresses. In addition to meeting these requirements of tire service, the cord fabric must possess characteristics permitting of satisfactory processing which may include passage through an aqueous rubber dispersion, drying at an elevated temperature, calendering, tire building, tire expansion from approximately cylindrical to toroidal form, and finally curing in a mold at high temperature.

The foregoing brief recital of the requirements of tire cord fabric indicates the importance of properly evaluating the cord materials from the standpoint of processing and tire service. This evaluation falls naturally into three broad classifications: (1) fundamental chemical and physical studies of the base materials, (2) investigations of the properties of the various materials in cord form in relation to the various physical parameters involved in tire processing and service, and (3) determination of the optimum cord construction and treatment for each material. The discussion which follows will be concerned mainly with two parameters of the second classification: temperature and moisture.

PART I. TENSILE PROPERTIES

Significance of the Tensile Properties

Tenacity and elongation, as measured with a conventional tensile testing machine, have very little direct relation to the serviceability of a tire cord. A cord of high tenacity may prove to be low in dynamic fatigue resistance; a cord showing a low "10-pound stretch" may be found to creep seriously under dead load. Still, in the majority of cases, the tensile properties may be considered as fairly reliable quality indices. Familiarity of textile and rubber technologists with measurements of this type and the simplicity of the measurements justify their employment for basic studies of cord behavior, providing no final conclusions be drawn until creep and fatigue results have been obtained. Hence, it appeared desirable to begin our studies of the effects of temperature and moisture with tensile measurements over a wide range of temperatures and relative humidities.

The mechanism of sorption of water by cellulosic materials has been widely studied. An excellent summary of the subject is given by Valko.[1]

The effects of humidity and temperature on the tensile properties of individual textile fibers and filaments have received considerable attention.[2-5] Wiegerink[6-8] devised equipment with which he determined absorbed and desorbed moisture of several types of textile yarns after treatments at various elevated temperatures and relative humidities. He found that the absorbed and desorbed water contents of variously treated

[1] E. I. Valko, *Cellulose and Cellulose Derivatives* (Interscience Publishers, Inc., New York, 1943), edited by Emil Ott, pp. 379–423.

[2] Brown, Mann, and Pierce, J. Text. Inst. **21**, T187 (1930).

[3] H. Hindman, and K. R. Fox, Rayon Text. Monthly **14**, 43 (1943).

[4] H. Leaderman, Elastic and creep properties of filamentous materials and other high polymers, The Textile Foundation, Washington, D. C. (1943).

[5] W. Whitehead, Rayon Text. **24**, 456 (1943).

[6] J. G. Wiegerink, J. Research Nat. Bur. Stand. **24**, 639 (1940).

[7] J. G. Wiegerink, J. Research Nat. Bur. Stand. **24**, 645 (1940).

[8] J. G. Wiegerink, J. Research Nat. Bur. Stand. **25**, 435 (1940).

cottons, and viscose, cuprammonium, and acetate rayons decreased with increasing temperature in the range 35.5 to 150°C. He also measured changes in the tensile properties of the yarns resulting from extended conditioning at various temperatures and humidities but the tensile measurements were made at room temperature. His results constitute a definite contribution to knowledge of the permanent effects of humidity and moisture treatments. It should be remarked, however, that Wiegerink's definition of relative humidity[8] at temperatures above 100°C appears to be incorrect. He defines relative humidity as "the ratio of the actual pressure of water vapor to the maximum possible pressure of water vapor in the atmosphere at the same temperature, expressed as a percentage." Since his conditioning chamber was at atmospheric pressure, this definition is meaningless. Busse, Lessig, Loughborough, and Larrick[9] give curves relating the "tensile strength" of cotton and rayon tire cords to temperature, measured under "oven dry" conditions but give no data obtained at elevated humidities.

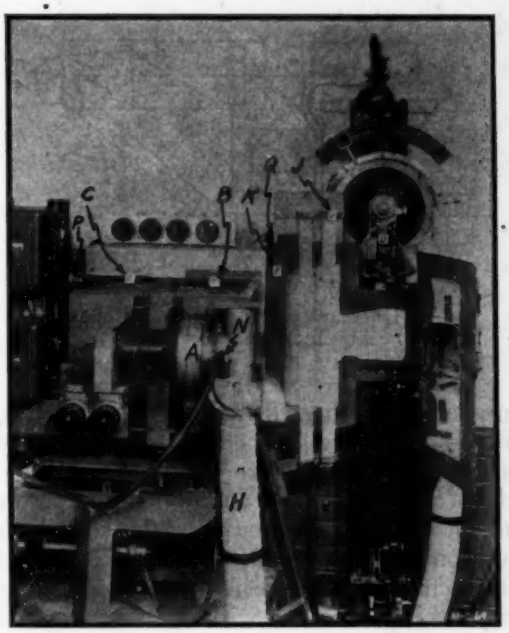


FIG. 2. Cord humidifier—cord loading door open.

Apparatus

A standard type X-3 Scott Tensile Machine, mounted in a room conditioned at 21.1°C and 65 percent R. H. was employed. A special conditioning unit, shown in Figs. 1 and 2, was so constructed that it could be wheeled into the fabric laboratory and employed to condition the cords before and during the tests, without moving the tensile machine. With the grips adjusted to test 15.4-inch cord lengths, the testing chamber *G* just fit between them. Groups of 30 cords were held in clamps *J* and conditioned in the wide chamber *F* for 1 hour before being pulled along the felt-lined slot *K* to the test position. Three groups of cords could be conditioned simultaneously. A close-up of the machine with door open is shown in Fig. 2. When a cord was actually being tested, it rode in a vertical groove, the felt lining of which barely touched the cord. The full 15.4-inch gauge length of cord was conditioned in *F* but, of course, as each cord was brought to the test position and clamped around the standard grips, the ends were exposed to the conditioned room atmosphere. Then, as the test progressed (12 inches/min.), more of it was exposed at the lower end. Thus values of elongation were obtained on a cord which was not entirely enclosed and, therefore, not strictly at the chamber temperature. However, since breaks always occurred inside the chamber, the observed breaking strengths were characteristic of the chamber temperature and humidity. Data on both tenacity and elongation are presented but it must be remembered that the latter are only approximately correct.

The method for achieving the desired temperature and humidity in the chamber is best understood by referring to the schematic top view of Fig. 3 as well as Figs. 1 and 2. Air circulation was maintained by a motor-driven blower *A*. Steam injected into the air stream at *N* saturated the air with water vapor in a baffled compartment *B*. The hot saturated air then passed into a second chamber *C* where it circulated through a radiator core (from an automobile heater) which reduced the air temperature to that desired, producing saturated air at a temperature T_P which was measured at point *P*. The radiator core was maintained at the desired temperature by means of a pump-circulated water system, in which was immersed a knife heater, the current for which was controlled with a bimetallic spiral-Mercoid relay assembly. The spiral was located at point *P*. In order to obtain regulation at lower temperatures, a cooling coil was also provided in the water system. The saturated air at

[9] Busse, Lessig, Loughborough, and Larrick, J. App. Phys. **13**, 715 (1942).

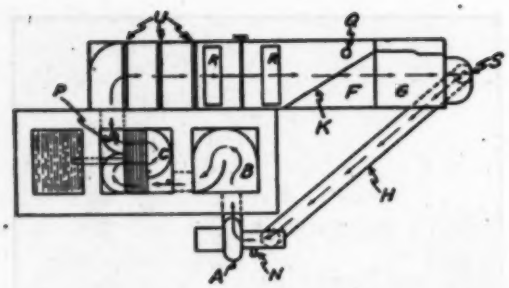


FIG. 3. Schematic top view of cord humidifier showing circulation path.

temperature T_P passed next through a series of Nichrome heating elements U where it was heated to a temperature T_Q and stirred with paddle agitators R to minimize temperature gradients before it passed over the cords in the conditioning chamber F and the cord in the testing chamber G. The temperature T_Q of the air was kept constant with another bimetallic spiral Q and Mercoid control assembly which varied the current in the Nichrome heating elements. The air finally returned to the blower through the flexible tube H. The duct system was made of Transite to as large an extent as possible to minimize heat losses and the tube H was wrapped with asbestos tape.

The temperature T_K at the cord conditioning location K was determined by means of a calibration curve giving T_K (as measured by a fine thermocouple) in terms of the temperature T_S at point S. Thus it was unnecessary to employ thermocouples at K except when performing the calibrations. The mean temperature T_G of the box at the test position was slightly lower (3 to 5°C) than T_K but, because of the short time during which the cord was at the lower temperature before the break, this small temperature gradient was ignored.

A dew-point psychrometer used for periodic checks of the moisture content of the air was also located at point S. A small window (not shown in Fig. 1) was installed to permit observation of the polished nickel tube of the psychrometer. The tube was insulated with asbestos tape to within one inch of its lower end. The top of the tube was closed with a rubber stopper, in which were inserted a thermometer, a short glass tube extending just inside, and a long glass tube extending to the bottom of the nickel tube. The psychrometer was partially filled with ether. The short glass tube was connected to an aspirator, causing air to bubble through the ether and lower the temperature. The dew point was read by means of the thermometer immersed in the ether.

The relative humidity for any test temperature T_K was defined simply as the ratio of the density of water vapor at point K to the density of water vapor necessary to saturate the air at the conditions of point K, i.e., temperature T_K and atmospheric pressure. The density d_P of the water vapor present in the air emerging from the radiator core (at point P) was, of course, that corresponding to saturation at temperature T_P, as read from standard steam tables. As this air-water vapor mixture passed through the Nichrome heating elements, both the air and water vapor expanded according to Gay-Lussac's law. Thus, although the ratio of water vapor to air remained constant in passing from position P to the test position K, the air was not saturated at temperature T_K obtaining at position K and the absolute humidity at K was $d_P \cdot (T_P/T_K)$, T_P and T_K being expressed as absolute temperatures. For temperatures T_K below the boiling point of water (taken as 373°K), the density of water vapor for saturation D_K was that value given in standard steam tables, which may be denoted as D_K'. Thus, for T_K below 373°K, the relative humidity was:

$$\text{R. H.} = \frac{d_P \cdot (T_P/T_K)}{D_K'}. \tag{1}$$

For T_K above 373°K, however, it was necessary to reduce the density for saturation obtained from the steam tables D_K' to the value for atmospheric pressure by means of Boyle's law. Thus, D_K became:

$$D_K = D_K' \cdot (P_O/P_K),$$

where P_O is the atmospheric pressure and P_K is the pressure of saturated vapor at temperature T_K, as read from the steam tables. Thus, for T_K above 373°K, the relative humidity was calculated as:

$$\text{R. H.} = \frac{d_P \cdot (T_P/T_K)}{D_K' \cdot (P_O/P_K)}. \tag{2}$$

It should be noted that this calculation ignores departures from Boyle's law. However, these departures are quite small in this range of pressures. The calculation for T_K below 373°K is essentially the same as that employed in compiling dew-point tables. Unfortunately, such

tables seldom extend to temperatures above 100°C. The dew-point psychrometer at point S was employed for an additional measurement of relative humidity. The fact that values of relative humidity at point S, obtained from the psychrometer observations, agreed to within ± 2 percent with the corresponding values calculated by expressions (1) and (2) indicates that equilibrium had been reached at the test position K and that condensation of water vapor between points P and S was negligible.

The following test procedure was employed: Cords were kept in the standard conditioned room (21.1°C, 65 percent R. H.) for at least 24 hours before being inserted in the special conditioning unit. Three racks of 30 cords each could be mounted simultaneously in the conditioning position K. Tensile tests were generally made on 20 cords of each group, the remaining cords being available for moisture regain measurements and emergencies. Twenty minutes was required to test a set of 20 cords. Hence, racks of 30 cords were put into the conditioning unit at 20 minute intervals, each rack being allowed to remain in the conditioning position K for 1 hour before starting the test. Thus, the average conditioning period for the cords of a given rack was 70 minutes. This period was sufficient to ensure close approach to water absorption equilibrium but was not so great as to permit large permanent changes in the cord properties. The rate at which the moisture regain (percent moisture referred to "oven-dry" state) reached equilibrium is illustrated by Table I. The tensile measurements

TABLE I. Moisture regain of 1100/2 viscose rayon cord *vs.* time at 150°C, 60 percent R. H. (Cords preconditioned 3 months at 21.1°C, 65 percent R. H.)

Time at 150°C 60 percent R. H. (minutes)	Regain (percent)
30	2.25
60	2.37
90	2.17
150	2.20
210	2.21

were performed in the usual manner except that it was necessary to use a 15.4-inch length of cord instead of the standard 10-inch length, in order to accommodate a test chamber of reasonable dimensions between the grips. Ordinary oven-dry and conditioned tests in the open room were made with the 15.4-inch gauge lengths so that the results were comparable with the

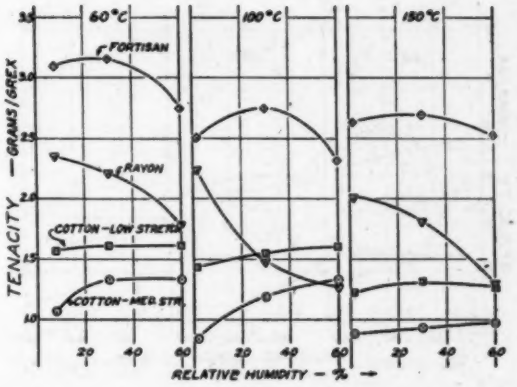


FIG. 4. Tenacity *vs.* relative humidity at constant temperature—cotton, rayon, and Fortisan cords.

tests made with cords in the special unit. It should be remarked that the relative humidities employed in the conditioning unit were 60 percent or lower and the temperatures were all above 21.1°C. The treatment in the box, therefore, always involved desorption since the preconditioning was at 21.1°C and 65 percent R. H. In most cases, a few cords were removed from the unit and quickly placed in weighing bottles previously oven-dried and weighed. The bottles were reweighed to obtain the wet weights of the cords and then dried with stoppers out at 120°C for 1½ hours. The regain was then calculated as the ratio of the loss in weight of the cord upon oven-drying to its oven-dry weight.

Experimental Results

Five different types of tire cord were chosen for this study: (1) 11/4/2 medium stretch cotton, (2) 11/4/2 low stretch cotton, (3) 1100/2 viscose rayon, (4) 1100/2 Fortisan, and (5) 210/3/3 Nylon. The resistance to rupture has been expressed as tenacity in grams/grex,* rather than as strength referred to unit area, in order to obtain a comparison of the various materials on an equal weight basis. The tenacity as a function of relative humidity at temperatures of 60, 100, and 150°C is given in Figs. 4 and 5. It is to be noted that the tenacity of the low stretch cotton cord increased much less with increasing humidity, as compared to the medium stretch cotton. This difference in behavior probably resulted mainly from the fact that the low stretch cord was much more compact and thus its tenacity was less influenced by the swelling

* The grex unit is defined as the weight in grams of 10,000 meters of cord. For a discussion of the Grex system, see A. G. Scroggie, Rayon Text. Monthly **25**, 45 (1944).

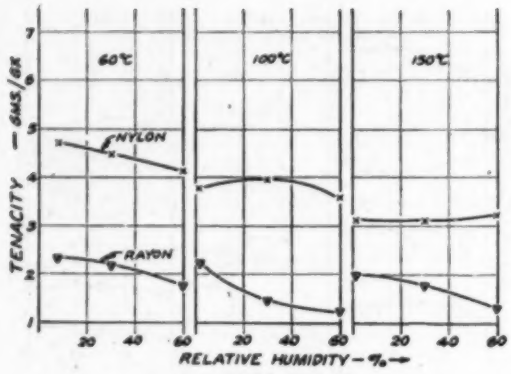


FIG. 5. Tenacity *vs.* relative humidity at constant temperature—rayon and Nylon cords.

of the cotton fibers. The low stretch cotton cord probably contained somewhat more wetting agent than the medium stretch cord which also would tend to explain the smaller humidity dependence of the former. This effect must be small, however, since it was found that the two types of cotton cord had essentially the same moisture regain. (See Fig. 12.) It should be remembered, of course, that the cotton fiber, itself, increases in tenacity with increasing humidity.[2] The commonly accepted but not too satisfactory explanation of this effect is that the swelling by water of the less organized cellulose in the intermicellar spaces results in an improved stress distribution and, therefore, higher tenacity at higher humidities. A similar explanation might be advanced for the smaller humidity dependence of the low stretch cotton cord for the process by which it is produced results in a more uniform cord. Hence the uniformizing action of absorbed water would be expected to have a smaller effect upon the tenacity of the more uniform low stretch cord. The available data do not appear to be adequate for a decision as to the relative importance of compactness and uniformity. The smaller increases in tenacity observed for both types of cotton cords at 150°C are undoubtedly the effect of the rapid decrease of moisture content between 100°C and 150°C. (See Fig. 12.)

The viscose rayon cord shows the expected decrease in tenacity with increasing humidity, resulting from the swelling of the individual filaments and consequent decrease in the intermolecular forces. The behavior of the Fortisan cord is interesting, a distinct tendency for a maximum tenacity at intermediate humidities in the region of 30 percent R. H. appearing. It might be said that Fortisan behaves like low stretch cotton at the lower humidities; like rayon at higher humidities. General considerations of the Fortisan process in which the cellulose is regenerated under stress from the acetate would seem to allow for a structure possessing both the discrete micellar nature of cotton and the continuous distribution of molecular weights characteristic of rayon. This rather reasonable speculation is certainly consistent with the observed behavior of the Fortisan cord. The over-all higher tenacity of Fortisan, as compared to rayon, might result from several different factors: higher state of orientation, higher degree of polymerization, etc.

The decrease of the tenacity of the Nylon cord at 60°C (see Fig. 5) is somewhat larger than might be expected from its low moisture absorption. It might be concluded that moisture has a large effect on its tenacity at this temperature, since, although a smaller amount of moisture is present, the total decrease in tenacity is about as large as in the case of rayon. At the higher temperatures (100 and 150°C), the effect is smaller for Nylon which is difficult to explain in the light of the smaller decrease of moisture content with increasing temperature found for nylon. (See Fig. 12.)

The tenacity of the cord as a function of temperature is given in Figs. 6 and 7. For the oven-dry condition (R. H. = 1.6 percent), all cords except Fortisan show the expected decrease of tenacity with increasing temperature. At relative humidities of 30 and 60 percent, however, the tenacity of the rayon cord increases with increasing temperature in the range 100 to 150°C. The slopes of the curves for the cotton cord increase greatly in this range also. Explana-

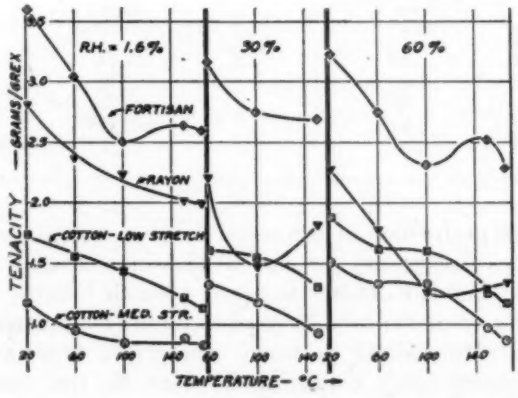


FIG. 6. Tenacity *vs.* temperature at constant relative humidity—cotton, rayon, and Fortisan cords.

tions of this phenomenon will be apparent later in connection with a discussion of Figs. 12–17.

The effects of humidity upon 10-pound stretch and the ultimate stretch are given in Figs. 8 and 9, respectively. In general, the extensibilities of the cords increased with increasing humidity, as is to be expected. The anomalous minimum in the 10-pound stretch of the medium stretch cotton at 30 percent R. H. and 100°C is difficult to explain. The stretch *vs.* humidity behavior of the Nylon cord is of interest since the dependence is negligible at 60°C but very large at 100 and

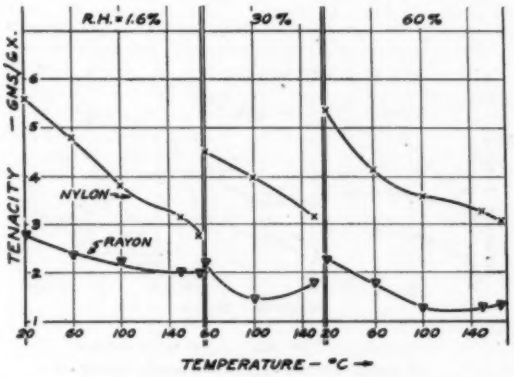


FIG. 7. Tenacity *vs.* temperature at constant relative humidity—rayon and Nylon cords.

150°C. It thus appears that the effect of moisture on the extensibility of Nylon is enhanced at the higher temperatures (see Fig. 9) while just the opposite behavior is found for its tenacity (see Fig. 5). The close parallelism of the stretch behavior of the Fortisan and the low stretch cotton cord shown in Figs. 8–11 is striking. Certainly, some of the characteristic properties of cotton have been reproduced in the Fortisan cord. Whether this combination of the charac-

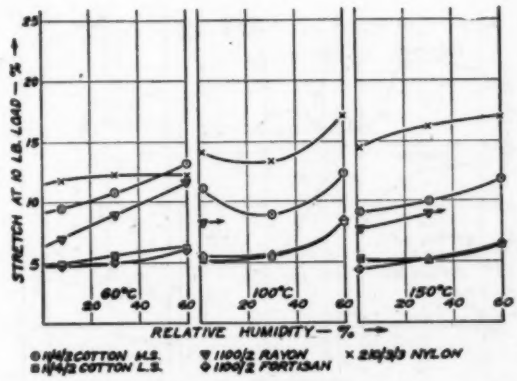


FIG. 8. Ten-pound stretch *vs.* relative humidity at constant temperature.

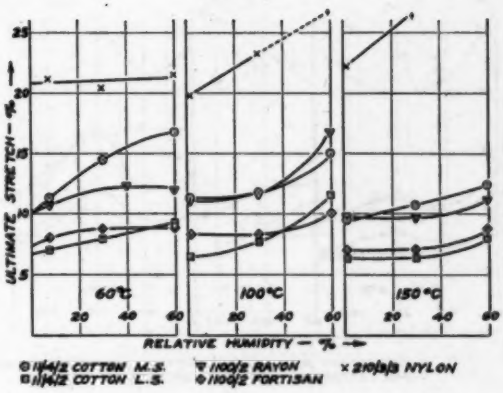


FIG. 9. Ultimate stretch *vs.* relative humidity at constant temperature.

teristics of rayon and cotton is desirable must be decided by other measurements which pertain more directly to tire service.

It should be emphasized that the temperature dependences of tenacity and extensibility at constant relative humidities of 30 and 60 percent (see Figs. 6, 7, 10, and 11) are quite misleading since the moisture content decreases rapidly with increasing temperature at a given relative

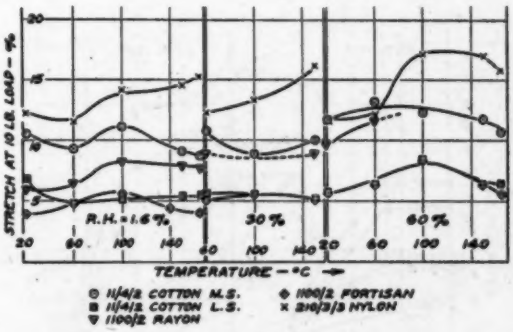


FIG. 10. Ten-pound stretch *vs.* temperature at constant relative humidity.

humidity. This fact is apparent from the curves of Fig. 12. Obviously, the correct method of studying the effect of temperature is to plot the tenacity and stretch against temperature for constant regain. In order to obtain constant regain curves of this type, it was necessary first to plot stretch and tenacity against regain (see Figs. 13 and 14) and then read the desired stretch or tenacity values at constant regain from the curves. An example of the much simplified picture which is obtained by plotting tenacity *vs.* temperature at constant regain for the rayon cord is given in Fig. 15. Here it is clear that

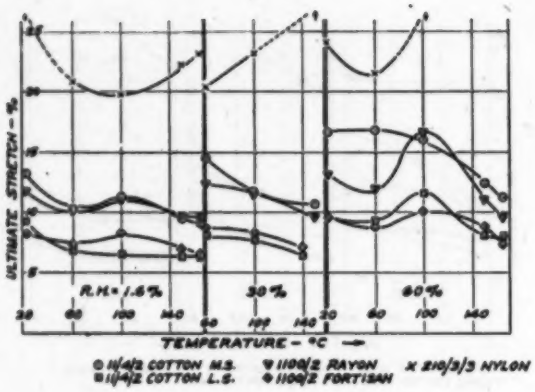


FIG. 11. Ultimate stretch *vs.* temperature at constant relative humidity.

rayon decreases in tenacity with increasing temperature, either for the oven-dry condition or for any given moisture content. This result is of interest since many misstatements concerning the effect of temperature on the tenacity of rayon have been made in the absence of data of this type. Certainly the well-established advantage of rayon over cotton cords in hot-running truck tires cannot be explained by these curves. However, they do suggest that rayon tires should

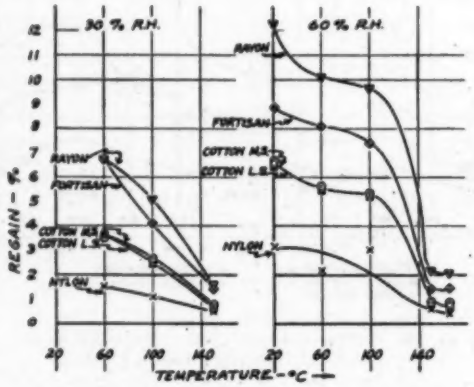


FIG. 12. Regain *vs.* temperature at constant humidity.

be built in a dry state in order to take advantage of the high dry tenacity of rayon. This principle is well established in the tire industry, after certain unfortunate experiences in the early history of rayon tire development where this precaution was not taken.

If a truck tire be built under conditions where the fabric is reasonably dry, there is every reason to suppose that the fabric will remain in a reasonably dry state since the rate of diffusion of moisture through rubber is quite low and, in operation, the tire will run at an elevated temperature which will counteract the inward diffusion process. The problem of a passenger tire, with its relatively low operating temperature, is somewhat different and it is quite possible that rayon might not show the advantage in a passenger tire, as found in a truck tire. This supposition is supported also by the fact that a smaller weight of rayon is generally used, as compared to the cotton cord found necessary in the same size tire. In a truck tire, this reduction of weight and the correspondingly reduced thickness constitute a great advantage in increasing

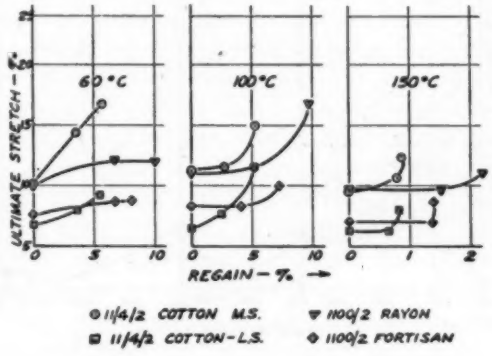


FIG. 13. Ultimate stretch *vs.* regain at constant temperature.

the rate of heat dissipation from the tire. An equivalent reduction in thickness of the passenger tire, where heat dissipation is a relatively small problem, offers only slight benefits. Thus, it may be necessary to use a greater weight of rayon cord in proportion to the cotton used in the same size passenger tire, than for the case of a truck tire, in order to realize the virtues of rayon.

Curves of tenacity as a function of tempera-

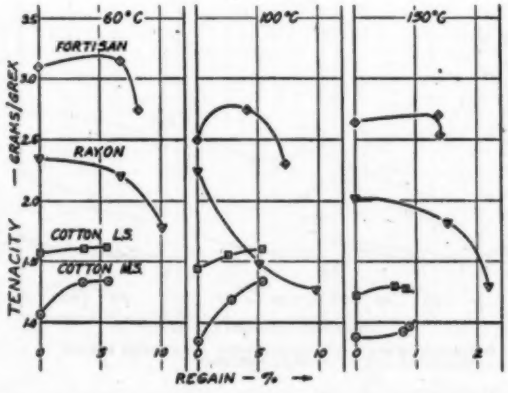


FIG. 14. Tenacity *vs.* regain at constant temperature.

ture at constant regain are given for the two types of cotton cords in Fig. 16 and for Fortisan in Fig. 17. Corresponding curves have not been plotted for Nylon because of the relatively small dependence of tenacity upon moisture content. The curves for cotton (Fig. 16) show quite different behavior for the low and medium stretch types: The tendency of the curves to flatten out at the higher temperatures, shown for the medium stretch cord, is absent in the case of the low stretch cords. The much smaller dependence of tenacity upon moisture regain shown by the low stretch cotton has been at-

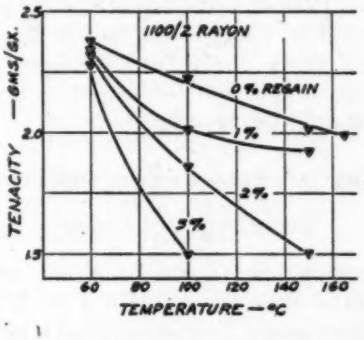


FIG. 15. Tenacity *vs.* temperature at constant regain—rayon.

tributed to its greater compactness, higher uniformity, and possibly the slightly greater wetting agent content. Regardless of which of these factors is the most important, the rapid decrease of the tenacity of the low stretch cord as the temperature rises above 100°C, cannot be considered favorable from the tire standpoint. Nevertheless, the higher tenacity of the low stretch cord as compared to the medium stretch cord, particularly for the oven-dry condition

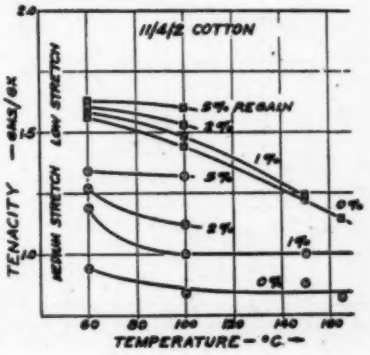


FIG. 16. Tenacity *vs.* temperature at constant regain—cotton

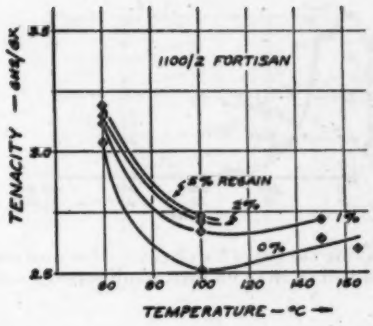


FIG. 17. Tenacity *vs.* temperature at constant regain—Fortisan.

(0 percent regain), at the higher temperatures is a point of superiority which is borne out by tire service tests.

The tenacity *vs.* temperature curves at constant regain for Fortisan show a minimum in the neighborhood of 100°C. It appears possible that the increase of tenacity from 100 to 150°C may be explained by thermal softening of the residual cellulose acetate (probably present in a very small amount) which might result in an equalization of the internal stresses. If the in-

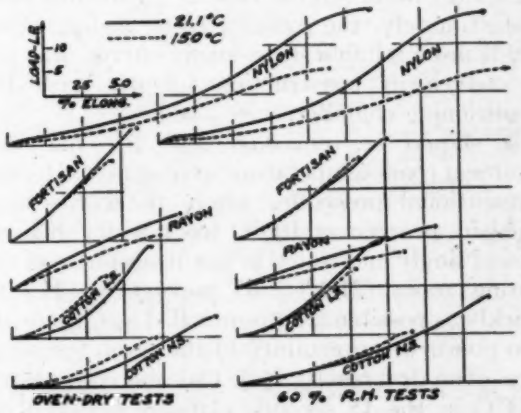


FIG. 18. Stress-strain curves from X-3 machine.

crease in tenacity in passing from 100 to 150°C be ignored, the shapes of the curves would resemble those found for medium stretch cotton. The moisture dependence of the Fortisan is in the same direction but of smaller magnitude than that of cotton. Thus, as mentioned earlier, the tensile properties of Fortisan appear to resemble those of cotton under some conditions; those of viscose rayon under others; and it appears possible that both the cotton and rayon types of structure may be present in Fortisan.

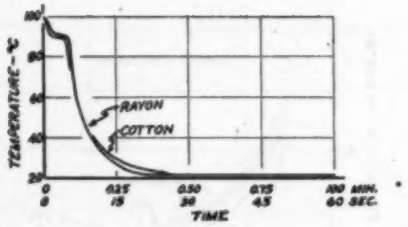


FIG. 19. Cooling curves of three-ply cotton and rayon cords removed from oven into conditioned room.

In order to indicate the changes in the stress-strain curves induced by variations in humidity and temperature, the curves of Fig. 18 were reproduced from the charts actually obtained in tests with the X-3 machine, corrections being made to obtain a roughly linear scale of ordinates. The dashed-line curves were obtained with cords at 150°C while the full line curves refer to room temperature (21.1°C) tests. Two sets of curves are given for oven-dry and 60 percent R. H. conditions, respectively. While the general behavior of the cord is summarized by these curves, it must be remembered that they do not accurately portray the stress-strain relation for reasons known to those who have employed the constant rate of stretch type of tensile machine. Unfortunately, the inclined plane tester, which yields more reliable stress-strain curves, was not of convenient construction for use with the conditioning chamber.

It should be remarked that the oven-dry results at room temperature were obtained by the conventional procedure where the cords were dried in an oven at 100°C for 1 hour; then removed singly and tested in the atmosphere of the testing room (21.1°C, 65 percent R. H.) as quickly as possible after removal. Thus, there are two points of uncertainty in the room temperature oven-dry results: (a) Did the cords reach 21.1°C in the 15 seconds elapsing between removal and break, and (b) was appreciable moisture absorbed in this period? In order to investigate point (a), fine (No. 40 B. and S.) iron-constantan thermocouple junctions were inserted at the centers of 3-ply cotton and rayon cords and the temperature *vs.* time curves during the oven-dry testing operations were obtained with a rapid recording potentiometer (Weston model 721). These curves are shown in Fig. 19. It is clear that the temperatures at the centers of the cords dropped to within 5°C of the testing room temperature at 15 seconds after removal from the drying oven. Thus, the break occurred with the cord at approximately 21.1°C. Nevertheless, during the early part of the stressing, the cords were at higher temperatures, which condition may have slightly influenced the breaking elongations and, therefore, the tenacities. The possibility (b) was studied by measuring the dependence of moisture regain upon time after removal from the oven (see Fig. 20). About 1.0 percent moisture was absorbed in the 15 seconds interval in the case of rayon and 0.5 percent for cotton. Thus, the room temperature oven-dry values of the tenacity were probably slightly high for the cotton and slightly low for rayon. However, these errors would be rather small so that the essential qualitative arguments drawn from Fig. 6 would not be invalidated. The presence of these uncertainties suggests, nevertheless, that a true oven-dry test at room temperature should be developed.

PART II. CREEP CHARACTERISTICS

Significance of Creep

Any segment of a cord in a tire in service bears a static inflation load of 1 to 3 lb. upon which is superposed a dynamic load which varies with the speed and the location of the cord segment in the tire. While the dynamic load undoubtedly affects the amount of creep of the cord to some extent, it appears reasonable to assume that the static load is the predominant factor. Because of the seriousness of the "growth" of tires over long periods of service, then, it was decided to make a study of the creep of the same cords employed in the tensile experiments of Part I, over a range of static loads and temperatures.

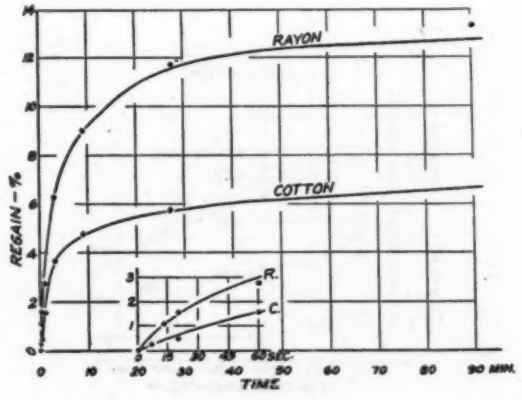


FIG. 20. Regain curves of three-ply cotton and rayon cords removed from oven into conditioned room.

Apparatus

The creep apparatus shown in Fig. 21 was essentially an insulated constant temperature box in which 20-inch lengths of cords were hung. the top grips were inside the box but the grips attached to the weight pans, which were prevented from rotation by means of guide studs, were just below and outside the box. A counterbalanced dial gauge, equipped with an accurately machined base, rested on the accurate stationary steel bar below the weight pans. Thus, the dial gauge could be moved successively to positions below each of the 12 loaded cords in the box and the vertical position of each weight could be measured as a function of time. Neglecting the short length of cord which emerged below the box as the test progressed, the elongation increment, defined as the additional elongation occurring after 0.002 hr., was calculated from the differences in dial gauge readings at time t and 0.002 hr., respectively. The initial length in all the calculations was taken as 20 inches under a "straightening load" of 1 oz. The origin of creep times was selected as 0.002 hr. for the reason that measurements seldom could be made at shorter times than 0.002 hr. after applying the loads. The cords were heated for 10 minutes after the thermometer reached the desired temperature, but before applying the loads, to insure uniform temperatures throughout the box and within the cords.

In order to obtain an approximate figure for the absolute elongation of each cord at 0.002 hr. after loading, an India ink mark was made at a point on the cord just outside the box, one minute after applying the 1-oz. load. The distance between the top grip and the bottom of the box was known and, hence, the initial length and the length at some later time when the creep rate was conveniently low could be measured. Comparison of these values with the corresponding dial gauge readings gave the absolute elongation where desired. It seemed preferable, however, to include the inherent inaccuracies of this type of measurement to as small an extent as possible in the creep results. Hence, the relatively more accurate elongation increment was used in plotting the creep curves and the absolute elongation at 0.002 hr. was calculated as a separate index to express approximately the extent of the rapid deformation occurring previous to that time.

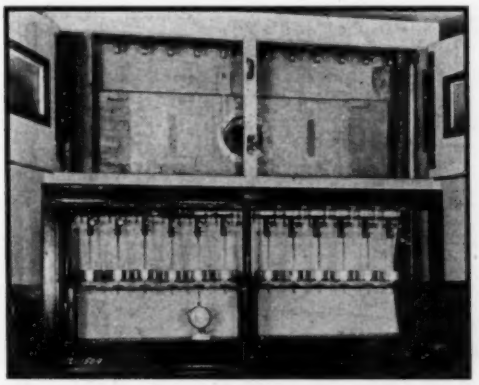


FIG. 21. Cord creep apparatus.

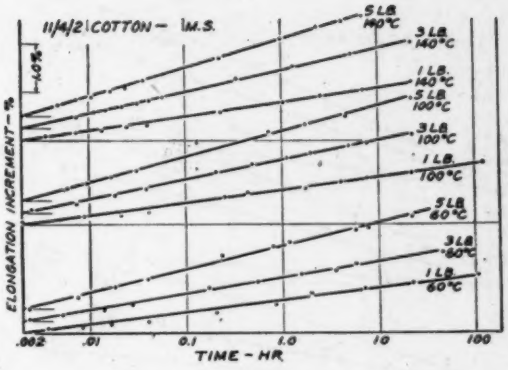


FIG. 22. Elongation increment *vs.* logarithmic time—cotton *MS*.

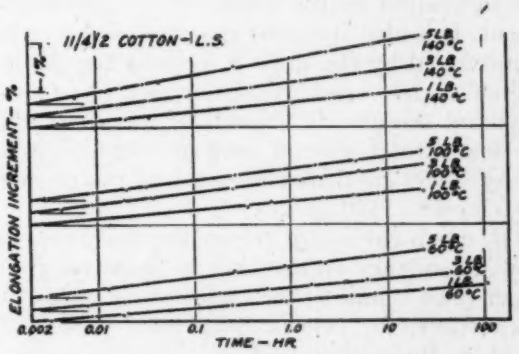


FIG. 23. Elongation increment *vs.* logarithmic time—cotton *LS*.

Experimental Results

The elongation increment is plotted against logarithmic time in Figs. 22–26. The creep curves for the two types of cotton cords (Figs. 22

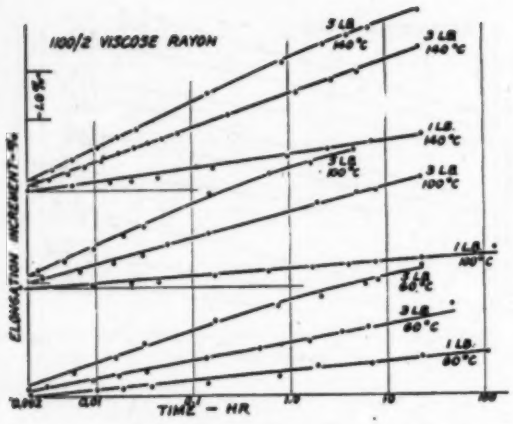


FIG. 24. Elongation increment *vs.* logarithmic time—rayon.

and 23) are seen to be linear in the time period investigated. The dependence of the slope S upon load is seen to be of the type expected, although S does not increase proportionally to load. The slope increases with increasing temperature, suggesting that creep takes place within the individual cotton fibers to a considerable extent. It is also apparent that the rate of creep, as measured by the slope S, is lower for the low stretch cotton cord, consistent with its lower 10-pound stretch. It should be remarked that the low stretch cotton cord was considerably lighter (4950 grex) as compared to the medium stretch cord (5580 grex).

The creep curves for the rayon cord (Fig. 24) show a tendency to decrease in slope for times longer than 1 hour at the highest load (5 lb.). In this respect, the curves differ from those published by Busse *et al.*,[9] where a distinct increase in slope was shown. However, the cords employed in that work were of different construction (275/4/3) and most of the curves presented were for temperatures of 150°C or higher. Hence, the two sets of data are not strictly comparable. At the higher loads, the creep of the rayon cord (Fig. 24), was much greater than that for either of the cotton cords. On the other hand, the rayon cord was much lighter (2660 grex) than the low stretch cotton cord (4950 grex). Thus, the creep at 3-lb. load for the rayon cords may be roughly compared with that at 5-lb. load for the low stretch cotton. On this equal weight basis, the rayon cord still appears to creep more rapidly than the low stretch cotton, in keeping with the known greater growth of rayon tires. To make a more exact comparison, of course, the creep loads chosen for cotton and rayon, respectively, should be equal to the inflation loads for individual cords in tires, as built (not on an equal weight basis).

The Fortisan cord (2920 grex) appears to creep considerably less than rayon (2660 grex) at the higher temperatures and loads and a slight tendency for the slope to increase is apparent (Fig. 25). This tendency is much greater for Nylon (Fig. 26) which shows considerable departures from linearity suggesting that chemical degradation takes place at the higher temperatures. Nylon also exhibits anomalous behavior as regards the dependences of creep upon load and temperature. For Nylon, and possibly to a lesser extent for the other materials, it appears that the initial elongation e_0 must be taken into account as well as the slope S.

In order to compare both the initial elongation and the slope of the creep curves on an equal weight basis, two arbitrary indices were defined: initial compliance $= e_0/P =$ elongation at 0.002 hr. divided by the unit load P in grams/grex; weighted creep $= S/P =$ slope of logarithmic creep curve divided by the unit load P in grams/grex. It should be noted that the slope S is the change in the elongation increment in the period t_1 to t_2 divided by $\log (t_2/t_1)$.

These indices were calculated at the 5-lb. load for the two cotton cords and at the 3-lb. load for the others, in order that the comparison would be at approximately the same "unit load" (load/lineal weight of cord). This procedure was desirable since the slope *vs.* load relation was not linear. The data are summarized in this form in Table II at temperatures of 100 and 140°C. The 60°C data were omitted since they were subject to errors from moisture effects. As previously judged, from casual examination of the

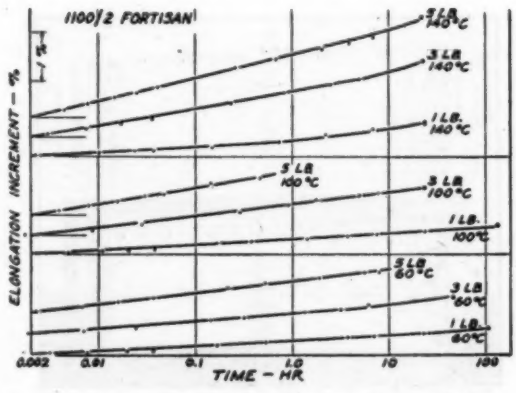


FIG. 25. Elongation increment *vs.* logarithmic time—Fortisan.

creep curves, rayon appears to have a much higher creep than that of low stretch cotton, although the initial compliance is much lower. Fortisan appears comparable to low stretch cotton in creep but much stiffer from the viewpoint of initial compliance. Nylon has about the same creep as that of rayon but is much softer in initial compliance. It must be remembered, of course, that the figures for the slope S are based on the early linear portions of the creep curves. Thus, the values of S/P for Nylon given in Table II, must be considered minimum figures, for the slope S increases at the longer times of creep. Because of the non-linear dependence of S upon load and the uncertainty concerning the constancy of S at extended creep times, it is clear that these laboratory creep measurements must be considered only as qualitative indications of the creep characteristics, as related to tire growth. It is interesting that the creep curves for cords of viscose rayon and Nylon are of same general type found for single filaments of viscose rayon and Nylon by Leaderman.[4]

TABLE II. Weighted creep S/P and initial compliance e_0/P.

Temp. (°C)	Index	M.S. cotton	L.S. cotton	Rayon	Fortisan	Nylon
100	e_0/P	20.8	14.1	2.9	5.1	10.4
	S/P	1.3	0.52	1.0	0.48	1.0
140	e_0/P	24.6	14.3	5.0	5.6	14.0
	S/P	1.5	0.74	1.3	0.71	1.5

Note: Both indices expressed as percent grex/gram.

SUMMARY

It may be concluded from the foregoing experiments that measurements of tenacity and stretch at a single relative humidity have little meaning since the important variable, moisture content, decreases rapidly with increasing temperature. The most revealing representation of tenacity or elongation appears to be as a function of temperature at constant moisture regain. Curves of this type show that the tenacity of a modern viscose rayon cord decreases with increasing temperature in the range of regains 0–5 percent. The same general behavior is exhibited by low stretch cotton cord but the tenacity *vs.* temperature curves for medium stretch cotton cord flatten off at temperatures above 100°C. The tenacity of Fortisan cord decreases rapidly with temperature up to 100°C, and then increases slightly between 100° and 165°C. Nylon cord shows a rapid decrease of tenacity with increasing temperature but maintains its advantage in tenacity over that of rayon for all temperatures and humidities included in these experiments.

For any given temperature, increase of moisture content increases the tenacity of the cotton cords to an extent dependent upon the "stretch" of the cord; decreases the tenacity of rayon cord. Fortisan increases in tenacity with increasing moisture content in the lower range of moisture contents (like cotton) and then decreases in tenacity for higher moisture contents (like rayon). The tenacity of nylon decreases very slightly with increasing moisture content.

All of the cord materials investigated, cotton, rayon, Fortisan, and Nylon, increase in stretch (10-lb. or ultimate) as the moisture content increases. The effect of temperature at constant regain upon stretch, however, is rather complicated. Until a method of measuring stretch with the whole cord continuously within the testing chamber is devised, further conclusions would be unjustified.

Analyses of the creep characteristics of the various cord materials on the basis of two arbitrarily defined indices "initial compliance" and "weighted creep" yield several basic facts: The low stretch cotton cord exhibits lower values of both indices as compared to medium stretch cotton, as might be anticipated from tensile measurements. The rayon cord, however, has a much lower "initial compliance" but a definitely higher "weighted creep," as compared to the low stretch cotton. The Fortisan cord is comparable in "weighted creep" to low stretch cotton but has a much lower "initial compliance." The Nylon cord has an "initial compliance" approach-

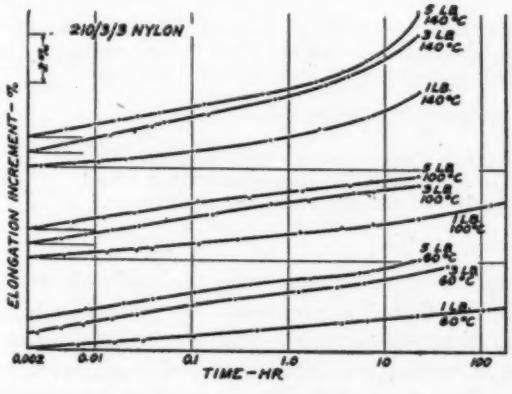


FIG. 26. Elongation increment *vs.* logarithmic time—Nylon.

ing that of low stretch cotton but a much higher "weighted creep" which is equal to or greater than that of rayon. The elongation increment *vs.* logarithmic time curves for cotton, rayon, and Fortisan are very nearly linear over the period 0.002–20 hr. but those for Nylon show a tendency to increase in slope at times beyond 1 hour.

In conclusion, it may be said that the results of these experiments agree very well with current concepts of the structures of cotton, rayon, and Nylon and have rather interesting implications in regard to the structure of Fortisan. It is well to remember that a complete evaluation of a tire cord should include dynamic fatigue measurements as well as tensile and creep data. The complicated nature of the fatigue problem, however, necessitates an extended separate discussion which is beyond the scope of this paper.

ACKNOWLEDGMENT

The authors wish to thank J. W. Liska who conducted some of the early experiments with the Conditioning Apparatus and F. S. Grover who built the equipment. We are indebted to L. L. Williams for his careful supervision of the tensile tests. The interest of J. N. Street in the results of the work and the permission of the Firestone Tire and Rubber Company to publish them is appreciated.

Numerical Solution of Ordinary and Partial Differential Equations by Means of Equivalent Circuits

GABRIEL KRON

General Electric Company, Schenectady, New York

(Received October 12, 1944)

Numerical methods are developed to solve certain types of linear and nonlinear partial differential equations to any desired degree of accuracy with the aid of equivalent electrical networks. The methods of solution of ordinary differential equations, both linear and nonlinear, follow as special cases. Three types of problems are considered:

1. *Initial-value problems.* If the field quantities are known along a surface, the networks may be solved by a straight-forward step-by-step calculation. The networks may also be looked upon as supplying a "schedule" of operations that can be put on a digital calculating machine.

For time-varying problems new types of networks are developed in which time appears as an extra spatial variable. Examples of new networks for the elastic field and for the general nonlinear wave equation are given. Sample calculations and theoretical checks of a transient heat flow problem and of an ordinary differential equation are also included.

2. *Boundary-value problems.* Four methods of solution are given, the first three being cut and try processes. (a) The method of weighted averages; (b) The method of unbalanced currents and voltages; (c) The "relaxation" method; (d) The "diffusion" method, that changes the boundary-value problem into an initial-value problem by adding to the original partial differential equation a time variable of the form $A\partial\varphi/\partial t$, allowing the unbalanced currents to "diffuse" in time.

These numerical methods may also be used to improve the accuracy of the results found on the Network Analyzer. Examples of calculations are given for the electromagnetic and the elastic fields.

3. *Characteristic-value problems.* Their methods of solution are similar to those of boundary-value problems. An additional method of unbalanced admittances is also indicated. It is shown that by calculating the power in the network, the characteristic value of the assumed function is found.

An improved characteristic value of the linear harmonic oscillator, solved initially on a Network Analyzer, is calculated as an example.

In general the electrical networks may be used to check the consistency and correctness of solutions arrived at by other methods, approximate or exact. The unbalanced currents at the junctions (easily calculated) give a quantitative measure of the deviation from the correct solution.

NETS

THE method of solving partial differential equations numerically by the use of equivalent circuits is by no means new. It has long been known[1–2] that Laplace equation $\nabla^2\varphi=0$ in rectangular coordinates may be solved to any desired degree of accuracy by (considering the two-dimensional case) drawing a net of squares, Fig. 1, each with area $\Delta x\Delta y$, and associating with each line the real positive number, unity. If with each junction (representing a point of space) the

correct value of the dependent variable φ is associated, then each φ must be the average of the four neighboring φ's. An extension of this method to Poisson's equation $\nabla^2\varphi = k$, where k is an arbitrary function, has also been in use.

The equivalent circuits presented here may be looked upon as generalizations to more general partial differential equations of the type of net of Fig. 1*a* used by mathematicians to represent Laplace's equation. In particular (considering Fig. 1*b*):

1. The square net with two lines crossing at right angles is replaced by nets with more complex configurations. Several lines may now meet at a junction and at arbitrary angles.

2. With each line is associated not unity but either a real, or an imaginary or a complex number a, b or $c+id$, positive or negative.

3. The junctions are associated not with one, but several dependent variables.

In these more general nets the variables represent line-, surface-, and volume integrals of field quantities on a set of interspersed small cubes into which the field is subdivided. Accordingly, not just any haphazard set of partial differential equations can be represented by a net. This physical interpretation requires that the equations should be in a tensor-density form.

It is a rather arbitrary but still valuable procedure to associate the numbers at the junctions by absolute potentials and the numbers along the lines by impedances, in particular (Fig. 2):
1. a positive real number by a positive resistor;
2. a negative real number by a negative resistor;
3. a positive imaginary number by an inductor;
4. a negative imaginary number by a capacitor.

This physical interpretation of the mathematical net allows additional physical concepts such as current, power, flux, etc., to be associated with the network, thereby simplifying its analysis and solution. It is true that in topology these additional concepts are also associated with the mathematical net, in particular current as "1-chain," power as "intersection-index," etc.,* but apparently these concepts have not yet been used for the solution of nets representing the Laplace-Poisson equations.

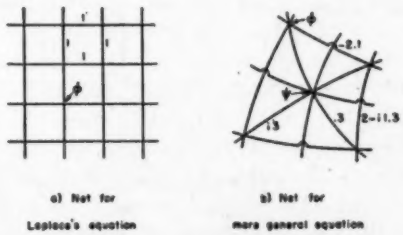


FIG. 1. Mathematical nets.

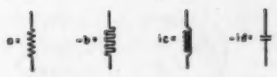


FIG. 2. Replacement of mathematical by electrical net.

In transient equivalent circuits not only numbers are associated with the lines of the net, but also operators, such as $\partial/\partial t$. In nonlinear circuits the line numbers are functions of the junction numbers. When only positive and negative real numbers exist, it is customary to replace a positive resistance by an inductance and a negative resistance by a capacitor (since none or only a few negative resistances exist on practical network analyzers).

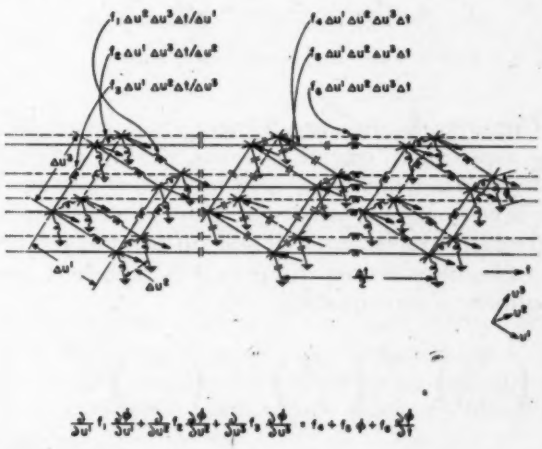


FIG. 3. Four-dimensional time-network.

Of course the algebraic symbols may be replaced by a mechanical model (masses and springs) or an acoustical model (pipes and chambers) instead of an electrical circuit.

PART I. INITIAL-VALUE PROBLEMS

Time-Dependent Equations

Partial differential equations containing time differ usually from others in the manner in which the boundary conditions are specified. In this part it will be assumed that at time $t=0$, the values of the dependent variables and their derivatives are known.

* See reference 1, p. 213.

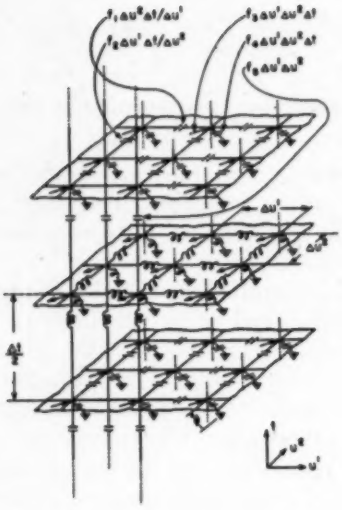


FIG. 4. Three-dimensional time-network.

Considering time as an extra spatial variable, the equivalent circuit consists of the original network (in which time does not enter) arranged in layers (one-, two-, or three-dimensional), separated by distances Δt and interconnected by suitable impedances. Figures 3–5 represent the nonlinear wave equation,

$$\frac{\partial}{\partial u^1}\left(f_1\frac{\partial\varphi}{\partial u^1}\right)+\frac{\partial}{\partial u^2}\left(f_2\frac{\partial\varphi}{\partial u^2}\right)+\frac{\partial}{\partial u^3}\left(f_3\frac{\partial\varphi}{\partial u^3}\right)$$

$$=f_4+f_5\varphi+f_6\frac{\partial\varphi}{\partial t},$$

for various numbers of independent variables.

In orthogonal curvilinear coordinates

$$f_1=\frac{h_2h_3}{h_1}f_1';\quad f_2=\frac{h_1h_3}{h_2}f_2'\quad f_3=\frac{h_1h_2}{h_3}f_3';$$

$$f_4=h_1h_2h_3f_4';\quad f_5=h_1h_2h_3f_5';\quad f_6=h_1h_2h_3f_6'$$

where the f'-s are arbitrary functions of u^1, u^2, u^3, and t.

When time enters as a second-order derivative, the networks are shown in Figs. 6 and 7. When time enters in a more complex form, different arrangements must be used.

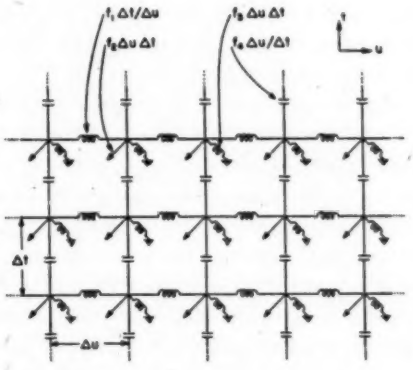
Figure 8 shows the time equivalent circuit of the elastic field. Since time enters as $\partial^2/\partial t^2$ for each dependent variable, all component networks are identical and for all spatial reference frames they are the same as those given in "Equivalent Circuits of the Elastic Field."[3]

It should be noted that while in the elastic and fluid flow fields the time coordinate may be left out or included at will, such is not the case in the electromagnetic field. In the latter the non-privileged role of the time coordinate necessitates a radically new treatment.

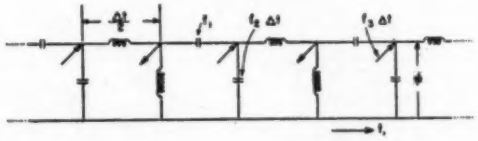
Time as a Spatial Variable

Let in a linear problem at $t=0$ an instantaneous impulse be impressed on the physical system. On the network this impulse is impressed by generators at $t=0$ and the voltage distribution gives the effect of the impulse in time and in space. (In the absence of a Network Analyzer the voltage distribution may be found by the methods to be shown presently.) Since the network extends to infinity along time, the network beyond any specified time t_0 may be replaced by a characteristic impedance.

If the effects of an impulse are known, then the effect of any other force applied within any



(a) Two-dimensional time-network.



(b) One-dimensional time-network.

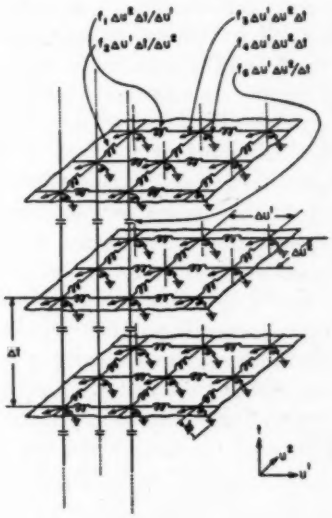
FIG. 5.

time interval may be found by simple superposition of the known voltage distribution.

It should be noted that the effect of more than one impulse cannot be studied simultaneously on the Network Analyzer, since a generator impressed, say, at $t=5\Delta t$ produces a voltage distribution at points corresponding to *previous* times, a situation impossible in classical dynamical problems. The numerical method to be shown prevents this "backsliding" of the impulses impressed at various times.

The Numerical Calculations

The numerical method is based on the fact that if all the potentials on the network are

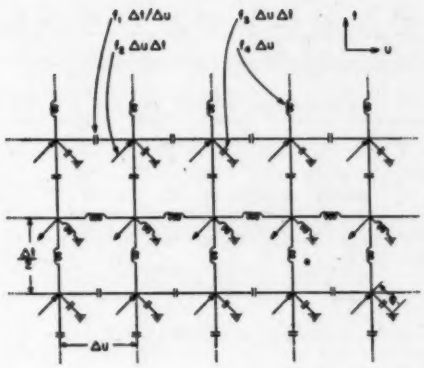


$$\frac{\partial}{\partial u^2} f_1 \frac{\partial \phi}{\partial u^2} + \frac{\partial}{\partial u^1} f_5 \frac{\partial \phi}{\partial u^1} + f_3 + f_4 \phi + \frac{\partial}{\partial t} f_6 \frac{\partial \phi}{\partial t}$$

FIG. 6. Three-dimensional time-network.

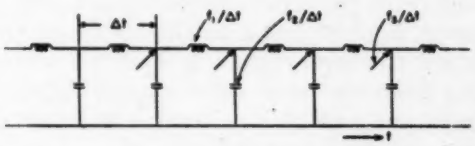
defined at $t=n\Delta t$ and at $t=(n+1)\Delta t$ (and also the boundary conditions along time), then all the potentials and currents can be calculated for the rest of the network in a straightforward manner up to any desired number of Δt's.

The numerical work consists of calculating all the admittances and currents on the *second* or $(n+1)\Delta t$ network. (Since in nonlinear problems the admittances are functions of the potentials, the former can all be calculated.) Adding up at a junction all the currents, including the one coming from the lower network, the resultant current must flow upward to the next network $(n+2)\Delta t$, Fig. 4. Since the admittance of the connecting coil is known (even in nonlinear



$$\frac{\partial}{\partial u} f_1 \frac{\partial \phi}{\partial u} + f_2 + f_3 \phi + f_4 \frac{\partial \phi}{\partial t}$$

(a) Two-dimensional time-network.



$$\frac{\partial}{\partial t} f_1 \frac{\partial \phi}{\partial t} + f_2 \phi + f_3 = 0$$

(b) One-dimensional time-network.

FIG. 7.

problems) the absolute potentials of the corresponding points on the next network $(n+2)\Delta t$ are thereby determined. The work of calculating currents is again repeated to proceed to the next network at $(n+3)\Delta t$.

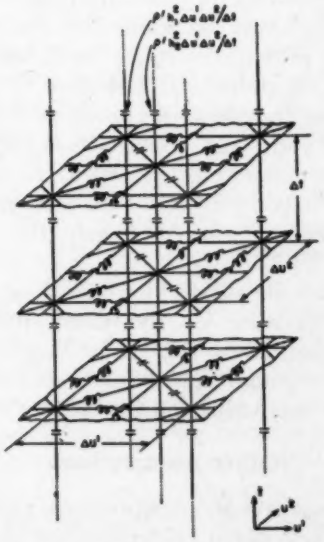


FIG. 8. Time-network of the elastic field.

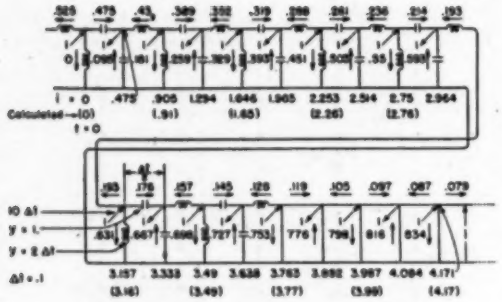


FIG. 9. Numerical example: $di/dt = 10 - 2i$.

Starting Assumptions

It should be noted that in order to get started, it is necessary to know not only all the currents on the network say at $t=0$ but also the currents coming from the next *lower* network, $t=-\Delta t$. However, in most problems the boundary conditions are so defined that at $t=0$ the currents coming from the lower network are not known, although the *sum* of the lower and upper connecting currents, namely $\Delta t(\partial\varphi/\partial t)$ is known at every point. *At the start of each problem thus the question arises how to divide the resultant currents into two parts at* $t=0$. Several methods are available:

(1) Half the current may be assumed to flow up, the other down. This assumes the slope of the curve $\varphi=f(t)$ to be constant in the neighborhood of $t=0$. After proceeding through two or three layers of networks it is possible to discern a more logical division than halving and to start all over again with the new divisions.

(2) Standard methods of approximations given in textbooks[4] may be used. For instance, after finding the potentials on the $t=\Delta t$ network, the differences of potentials between the two networks define the slope of the chord between the points $t=0$ and $t=\Delta t$. Using this slope instead of that of the tangent $\Delta t(\partial\varphi/\partial t)$ at $t=0$, a better division is accomplished. This step may be repeated several times between the first two networks.

(3) Δt may be subdivided into 5 or more subdivisions. Because of the finer divisions the results arrived at by the 5th Δt may be assumed to be close enough approximations and may be used as starting values on the coarser network.

Other Assumptions

Since the starting conditions are rarely known exactly, it will often be found that the currents or voltages in successive steps do not lie on smooth curves but oscillate around them in an attempt to correct themselves. Such overshooting and undershooting of smooth curves may be minimized by judicious changing at intervals of the division of currents that flow "up" and "down" between the layers. Most numerical methods do not possess this self-correcting property of the equivalent circuits.

When some of the admittances lying between two successive layers (between two Δt's) are functions of the dependent variables (hence, of time) their value has to be found by extrapolating from the known values of the dependent variables at two previous times.

Ordinary Differential Equations

In ordinary differential equations the independent variable may be either a spatial variable x or time t. If it happens to be the time, then usually sufficient information is available at $t=0$ to get the calculations started as above.

When the independent variable x is a spatial variable then usually only part of the boundary conditions (say the value of φ) are defined at $x=0$. Some boundary conditions (say the slope of φ) may, however, be defined only at $x=a$. In such cases it is possible to start out with a *tentative* slope at $x=0$ and find with the above step-by-step method the slope at $x=a$. Several new starts should eventually match the boundary conditions at the two different values of x.

An example,

$$\frac{di}{dt}=10-2i,$$

with initial conditions $i=0$ and $t=0$ (also $di/dt=10$ at $t=0$) is worked out in Fig. 9 in steps of $\Delta t=0.1$. The numerals in parenthesis give the results found in Doherty and Keller[5] for the same problem, found by a standard method requiring a cut and try procedure not only at the first step (between $t=0$ and $t=\Delta t$) but at every step.

Other worked out examples may be found in "The Application of Network Analysis to Some Electron Optical Problems."[13]

Example of a Transient Heat Flow Problem

Let the temperature at the edge of a plate be raised from zero to unit temperature at $t=0$ and be kept at that same temperature. The problem is to find the distribution of temperature along the plate (at distance x from the edge) at various times.

The equation of flow is

$$a\frac{\partial^2 T}{\partial x^2}=\frac{\partial T}{\partial t}.$$

Also $\Delta x=0.1$, $a\Delta t=0.0025$. Hence $\Delta t=0.0025/a$.

The boundary conditions are:

(1) at $t=0$, (*a*) at $x=0$ $T=1.000$

(*b*) at $0<x<1$ $T=0$

(2) at $t>0$, (*a*) at $x=0$ $T=1{,}000$

(*b*) at $x=1$ $\dfrac{\partial T}{\partial x}=0.$

Since the known data at $t=0$ are rather meager (T and $\partial T/\partial t$ are zero at every point along x except near $x=0$) for a preliminary calculation the assumed Δt is subdivided into five parts, so that $a\Delta t=0.0005$. The corresponding network to be used to get started is shown in Fig. 10.

The layer at $t=0$ only contains one current 5 flowing into the ground. (The existence of a ground is a boundary condition.) Hence this current cannot very well be assumed to divide into two equal parts. Instead, let it be assumed that a similar current in the next layer at $t=a\Delta t/2$ also has a value 5. Now half of that may flow up and half down. Assuming slightly less than half, say 2.45 flowing up, one potential on the next layer is 24.5. Its neighboring potential on the same layer is still undetermined but it must be small and could as well be assumed to be zero for the purpose of calculating the current flowing into this junction. The latter $(24.5-0)$ $0.005=0.1225$ is assumed again to flow half up, half down.

Continuing the calculation through five layers, the voltages on the fifth layer can be assumed to be approximately correct. Since the first layer contains too few informations, the calculation is continued to the tenth layer in order to be able to use the fifth and tenth layers as the first two layers of the coarser network of Fig. 11.

The calculations on Fig. 11 are straightforward. The numbers in parenthesis give the theoretically correct temperatures, showing satisfactory agreement with the approximate method. The farther the calculations are continued from the crudely approximated first layer, the better is the agreement with the analytical solution. It should be remembered that the set of vertical currents, by their deviation from a smooth curve, do show that some of the temperatures are too high and some are too low. Hence, they afford a means of correcting the results at intervals.

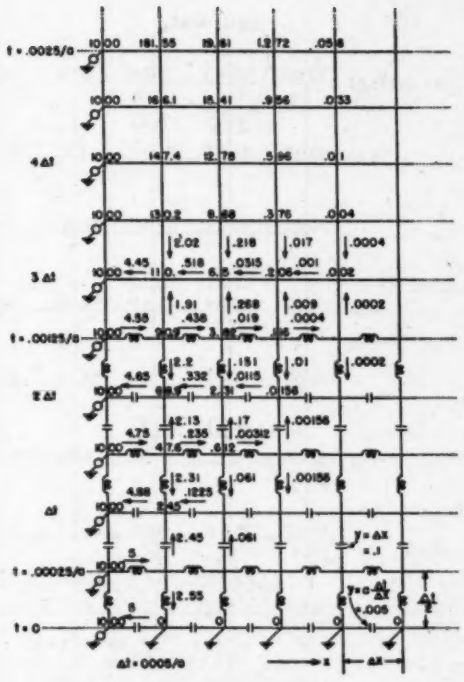


FIG. 10. Preliminary calculations.

Equations Not Containing Time

When an equation does not contain time, it still may happen that all field quantities are known upon a layer so that the next layer may be calculated. A class of such problems occurs in the design of magnetic or dielectric equipotential surfaces where a definite flux density (magnetic or electrostatic) is desired across a given surface. The problem is to shape the pole faces or electrodes to produce this desired flux distribution.

For instance, in a betatron[6] a known axially symmetric flux distribution is desired at the median plane of the accelerated electron. The equivalent circuit for the vector potential ψ_θ is shown in Fig. 12. Here the following quantities are known: 1. As all flux lines are perpendicular to the median plane, $H_{(r)}$ is zero and the vertical currents along the median plane are zero. 2. The flux distribution $B_{(x)}$ along the median plane are given, hence the differences of potentials along the corresponding coils are also known. 3. The current density in the current carrying region are known, hence all currents entering the junctions at the corresponding region are also known.

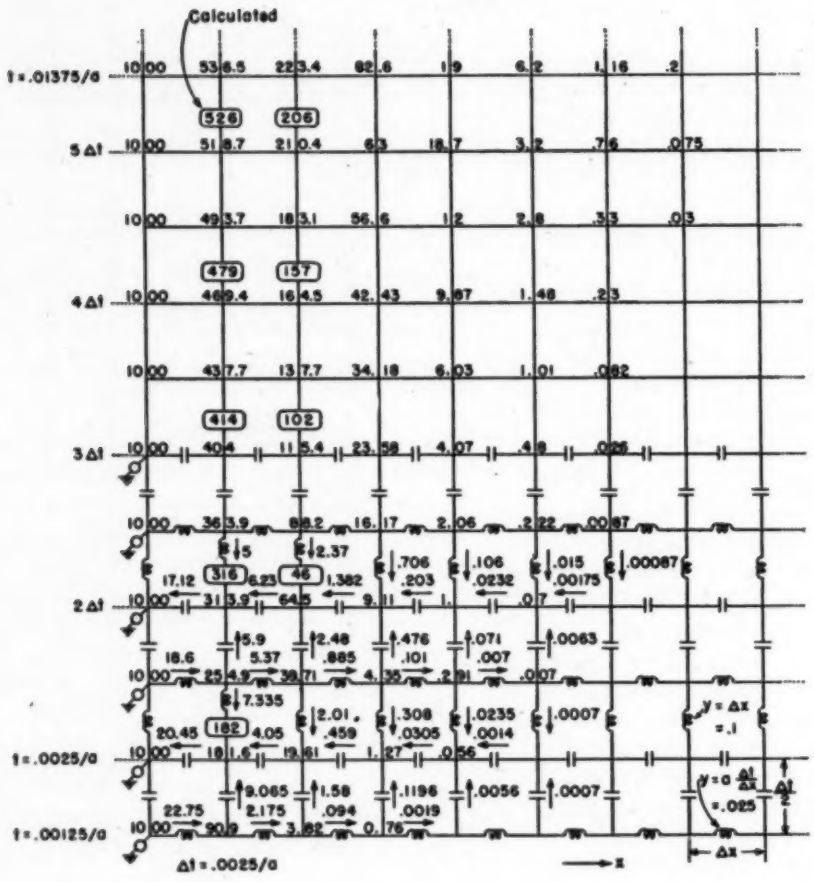


FIG. 11. Transient heat-flow calculations.

Sufficient data are available to continue the calculation of potentials and currents without any trial and error. The lines orthogonal to the equipotential lines represent the possible pole shapes to produce the desired flux distribution. When no current-carrying regions exist, a modification of the network, representing the scalar potential φ may be used, e.g., as shown in Fig. 10 of "Equivalent Circuits of Compressible and Incompressible Flow Fields."[7]

PART II. BOUNDARY VALUE PROBLEMS

The Problem

The equivalent circuits consist of a configuration of coils repeated along one, two, or three axes. The final shape of the network is the same as the shape of the field to be analyzed. While the illustrations will refer to two-dimensional networks, all rules and formulae are equally valid for one- and three-dimensional networks.

In analyzing a field, at first it should be divided into a very few subdivisions only. After its field distribution has been calculated, it should be subdivided into a finer mesh, then, if needed, into still finer meshes. In most problems it will be necessary to introduce fine meshes only in certain portions, usually around corners.

The problem in all cases is the following: (1) Given the boundary conditions in the form of known currents flowing in some of the coils and in the form of known absolute potentials at certain junctions. (2) Find the absolute potentials of all junctions and the currents flowing in all conductors.

In linear problems the impedances of all coils are also known. In non-linear problems only some of the impedances are known, while the value of others are functions of the as yet unknown absolute potentials and their derivatives. Some of the junction currents also may be functions of the potentials.

A Preliminary Step

In most networks the absolute potentials of the junctions have a physical meaning. In particular:

(1) In the electromagnetic field it is E_y, the field-intensity along the y-direction, or H_y. (2) In elasticity it is the total displacement u or v. (3) In hydrodynamics it may be the stream-function ψ or velocity potential φ, etc.

In all methods of attack as a preliminary step a tentative potential distribution at the junctions is assumed. Usually an intelligent guess may be made and if the guess is far off, that only entails more labor, without, however, influencing the correctness of the final answer. (Instead of the potential distribution, a guess may be made at the current-distribution.) Of course in linear problems this preliminary step may be done on a Network Analyzer, even though only a very few units are available, or the results of an approximate or exact solution may be used as starting points.

Once the absolute potentials at all the junctions are assumed to be known, it is always possible to calculate the currents flowing in all the coils. In non-linear problems at first the coil admittances and junction currents must be calculated from the known potentials, then the coil currents. Examples of non-linear networks are those of the potential flow of a compressible fluid in the physical plane, given in reference 7, Figs. 10 and 11.

The method of attack is based upon the fact that if the correct potential distribution is assumed and if all currents in the coils have been calculated then: (1) The sum of the currents

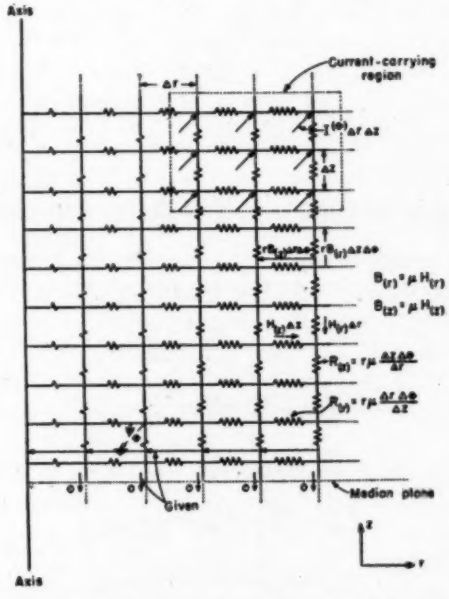


FIG. 12. Pole-face design of betatron (Magnetic vector potential ψ_θ).

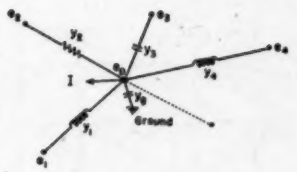


FIG. 13. A junction.

entering each junction is zero. (2) The weighted average of the potentials of neighboring junctions is equal to the potential of the central junction.

Three types of numerical calculations may be performed depending upon whether the weighted average of the currents entering a junction, or of the voltages surrounding a junction, or of both, are calculated.

Because of the bad guess, however, actually at no time will the above conditions be fulfilled at all junctions simultaneously. The purpose of the calculation is to reduce the difference in the entering and departing currents (or the difference between the surrounding and the central potentials) as close to zero as possible at all junctions.

This difference in the currents will be called the "unbalanced current" i_u, and the difference in the voltages the "unbalanced voltage" e_u. It should be noted that: (1) The "unbalanced voltages" offer a clue on how much to change the guessed-at voltages to reduce the error. (2) The "unbalanced currents" give an indication of how much the guessed-at values differ from the correct values.

Calculation of the "Unbalanced" Currents and Voltages

The unbalanced currents i_u at each junction are best calculated by summing up the currents entering it. To find the unbalanced voltage e_u at a junction whose absolute potential is e_0, let Fig. 13 be considered. The unbalanced current may also be found by

$$\begin{aligned} i_u &= (e_1-e_0)y_1+(e_2-e_0)y_2+\cdots(0-e_0)y_g+I, \\ i_u &= e_1y_1+e_2y_2+\cdots-e_0(y_1+y_2+\cdots+y_g)+I. \end{aligned} \tag{1}$$

If it is assumed that no unbalanced currents exist ($i_u=0$) then the correct value of the assumed potential e_0 would be E, the weighted average

$$\begin{aligned} i_u &= (e_1-E)y_1+(e_2-E)y_2+\cdots(0-E)y_g+I=0, \\ & e_1y_1+e_2y_2+\cdots-E(y_1+y_2+\cdots+y_g)+I=0. \end{aligned} \tag{2}$$

$$\textit{Weighted average} = E = \frac{e_1y_1+e_2y_2+\cdots+I}{y_1+y_2+\cdots y_g}.$$

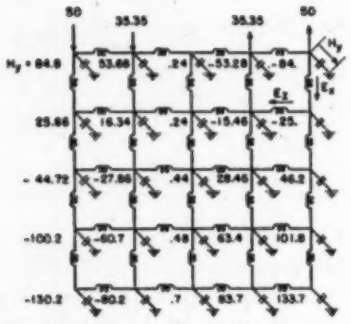


FIG. 14. Assumed potentials H_y on a square wave guide.

Since e_0 is assumed instead of E, the unbalanced voltage is

$$e_u = E - e_0, \tag{3}$$

$$e_u = \frac{e_1y_1 + e_2y_2 + \cdots + I - e_0(y_1 + y_2 + \cdots y_g)}{y_1 + y_2 + \cdots y_g},$$

$$e_u = \frac{i_u}{y_1 + y_2 + \cdots y_g}. \tag{4}$$

Hence the relation between the unbalanced i_u and e_u at every junction is

$$i_u = e_u(\sum y), \tag{5}$$

where

$$\sum y = y_1 + y_2 + \cdots + y_g \tag{6}$$

is the sum of the admittances of all the coils leading to the junction.

The Field Equations of Maxwell

In Laplace's equation, where $y_1 = y_2 = y_3 = y_4 = 1$ and $y_g = I = 0$,

$$i_u = e_1 + e_2 + e_3 + e_4 - 4e_0, \tag{7}$$

$$E = \frac{e_1 + e_2 + e_3 + e_4}{4}, \tag{8}$$

$$e_u = \frac{e_1 + e_2 + e_3 + e_4 - 4e_0}{4} = \frac{i_u}{4}. \tag{9}$$

In Maxwell's equation, in rectangular coordinates, Fig. 4 where $y_1 = y_2 = y_3 = y_4 = y_L$ and $y_g = y_c$ (a capacitance, a negative number),

$$i_u = [(e_1 + e_2 + e_3 + e_4) - e_0k]y_L, \tag{10}$$

$$E = \frac{e_1 + e_2 + e_3 + e_4}{k}, \tag{11}$$

$$e_u = \frac{e_1 + e_2 + e_3 + e_4 - e_0k}{k}, \tag{12}$$

$$e_u = \frac{i_u}{ky_L}, \tag{13}$$

where

$$k = 4 + y_c/y_L, \tag{14}$$

a number less than four.

That is, finding the unbalanced currents and voltages for the field equations of Maxwell in rectangular coordinates is no more complicated than finding the same quantities in Laplace's equation (i.e., the number 4 is replaced by k). It will be found in most problems that the general equations (1), (4), and (2) for finding i_u, e_u and E may be considerably simplified.

The Method of Unbalanced Currents and Voltages

One of the quickest ways (if not *the* quickest) to solve a complicated network is the following: 1. Assume (or measure by a Network Analyzer) the potentials at all junctions. 2. Calculate at all junctions the unbalanced currents i_u and the unbalanced voltages e_u. 3. Knowing the whole set of unbalanced quantities at the junctions assume a new set of potentials that are more correct. In general, *where e_u is positive, the potential should decrease and with negative e_u it should increase.* 4. Calculate over again the whole set of new unbalanced quantities at every junction.

Such successive guesses and calculations should reduce the unbalanced quantities below any desired amount. Plotting curves of the unbalanced currents, voltages, and of the guessed-at values at frequent intervals helps to speed up the convergence.

Examples of Unbalanced Quantity Calculations

Let a square wave guide or cavity resonator, Fig. 14, bounded in the x-z plane on three sides

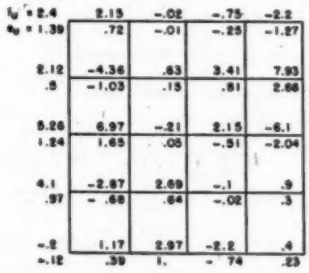


FIG. 15. Unbalanced currents and voltages on the wave guide network.

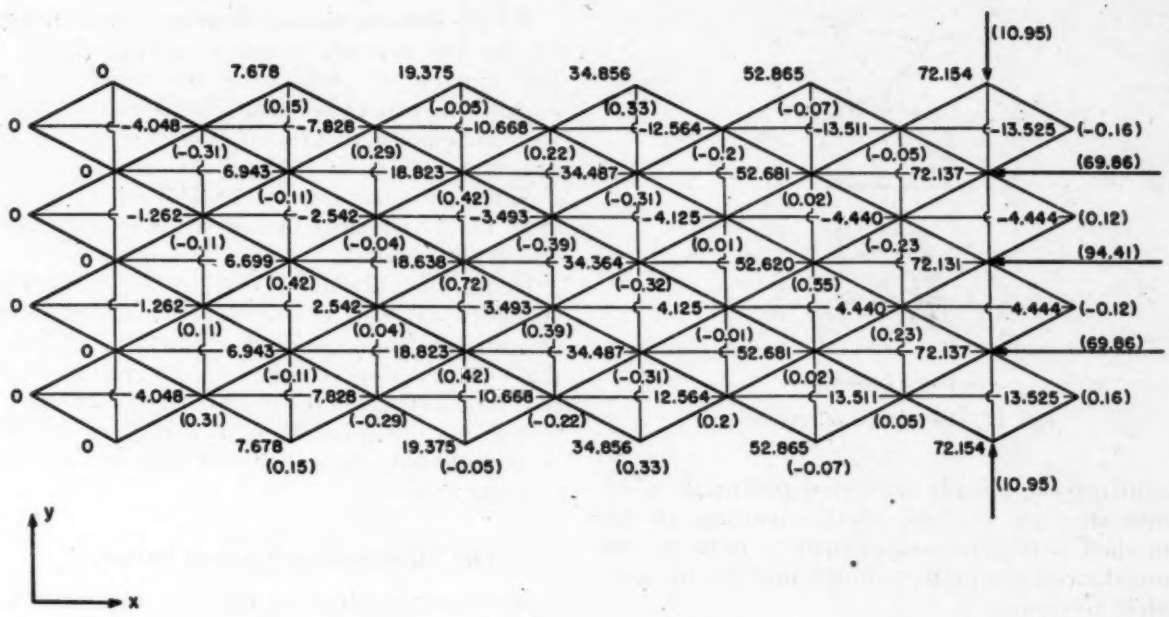


FIG. 16. Calculated displacements and the resulting unbalanced currents in an elastic beam problem.

by a conducting surface that extends to infinity in the y direction, be divided into sixteen squares. The net of coils in each square consists of two inductances ($z=0.8$ or $y=1.25$) and a ground capacitor ($z=1.3$ or $y=0.77$). The currents in the inductance represent E_x and E_z, while the voltages across the capacitors represent H_y.

For boundary conditions, it is assumed that at the open end (upper end) a sinusoidal current distribution in space, E_x, is impressed. As a result, a sinusoidal potential distribution H_y appears at the junctions. Neither the magnitude nor the wave-length of H_y is known.

Tentatively let the values shown on Fig. 14 be assumed as the junction potentials H_y. (Actually these values were found on the Network Analyzer.)

Figure 15 gives the unbalanced currents and voltages that will have to be liquidated. It should be noted that the greatest unbalanced junction current is 7.93 amperes. As the greatest coil current is 102.8 amperes, that represents an 8 percent inaccuracy.

Knowing the unbalanced currents and voltages at every junction, a new set of junction potentials may now be assumed.

For an elastic stress problem of a beam, a set of junction potentials (representing displacements in the x or y directions) and a set of unbalanced currents are shown in Fig. 16 (reproduced from Carter[8]).

The Method of Weighted Averages

Instead of attempting to reduce the unbalanced quantities simultaneously at all junctions, it is possible to reduce them at one junction at a time. Two such methods will be shown. (1) The unbalanced current and voltage are reduced to zero at a junction without compensating at the neighboring junctions (the method of weighted averages). (2) Each time the unbalanced current or voltage is reduced at a junction, the unbalanced quantities at the neighboring junctions are compensated simultaneously (the relaxation method).

The first method is especially valuable when the guessed at value of potentials differs from the correct value only at isolated points or regions. Such is the case usually in Network Analyzer solutions when at some junctions the instrument reading is incorrect or the board unit happened to be incorrectly set. The steps are as follows: (1) Assume (or measure) the potentials at all junctions. (2) Calculate E, the weighted average, at any junction by Eq. (2). (3) Replace e_0 by E.

In the previous method e_0 was not replaced by E. Only the corresponding e_u for each e_0 was calculated, leaving e_0 undisturbed.

It is customary to start at one corner of the network and change in succession all e_0 to E,

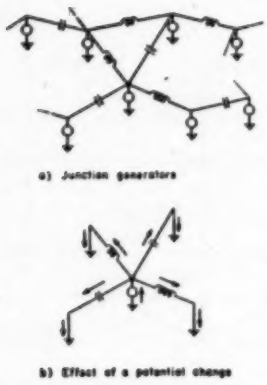


FIG. 17. The method of relaxation.

utilizing the already corrected potentials wherever they are available. A disadvantage of this method is that it is equivalent to reducing the unbalanced potential e_u at each junction immediately to zero.

A more advantageous procedure is to replace at each junction the value of e_0 not by E but by a value *larger* than E, or smaller (depending on the value of the neighboring potentials) *leaving thereby at each junction an unliquidated* e_u that is within the allowable limit of error.

In non-linear problems it is often possible to plot curves which give outright the value of E for given neighboring potentials (or rather for given potential-differences between neighboring junctions). If difficulties in the convergence arise, in place of E a value larger (or smaller) should be used.

The Relaxation Method

The relaxation method reduces the unbalanced currents or voltages at one junction at a time toward zero in a systematic manner. It is based upon the fact that whenever the absolute potential of a particular junction is changed, the unbalanced currents and voltages change only at that particular junction and at all the neighboring junctions to which a coil leads, while everywhere else the unbalanced junction currents remain unchanged.

This fact may be seen by introducing a set of hypothetical generators so that (Fig. 17*a*): (1) At every junction a generator with zero impedance is assumed to be connected to the ground. (2) The generator voltage is equal to the absolute potential of the junction. (3) The generator current to ground is equal to the unbalanced current existing at that junction.

By the introduction of these hypothetical generators the network remains unchanged and is now capable of performing new feats that it couldn't do otherwise. Since the generators have zero impedance to ground, *whenever one of the generator voltage is changed, currents can flow from this generator only to the neighboring grounds.* (Fig. 17*b*.) No change of currents exist anywhere else in the network.

Hence, the unbalanced currents are reduced in two steps: (1) Add (or subtract) a certain number of volts to the potential of a junction. (2) Find the new unbalanced currents (or voltages) at that junction and in the neighboring junctions.

Whether the unbalanced currents, or voltages, or both should be calculated depends on the problem at hand.

The "Unbalanced-Current Patterns"

The work involved in this last step may be greatly reduced in linear problems by the device of the "unbalanced-current pattern." This device consists of calculating once for all the *change* in the unbalanced junction currents caused by a change of *one* volt at every junction. (With non-linear coils this device cannot be used.) Two cases may be distinguished: (1) If the square nets of impedances throughout the network are identical, then each junction that does not lie on the boundary, has identical current-pattern. Similarly all horizontal, also all vertical, boundary junctions have identical pattern, as well as all corners. The various patterns for the waveguide example are given in Fig. 18 (*plus* current flows into a junction, *minus* current flows out of a junction). (2) If the impedances vary from point to point, then for every junction a different current-pattern has to be established.

In analogy to the "unbalanced current patterns" it is possible to set up "unbalanced voltage patterns" by multiplying each i_u by its respective $\sum y$.

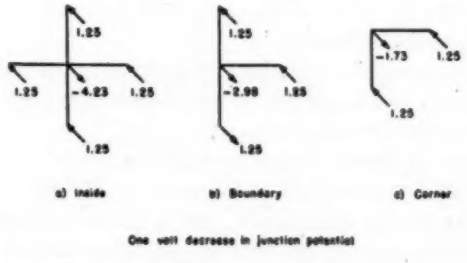


FIG. 18. Unbalanced current patterns.

The Numerical Reduction

It should be noted that every time a junction-voltage is decreased, the junction currents at the neighboring junctions increase, but at the junction itself it decreases. And the decrease is always greater than the increase (-4.23 amp. as against 1.25 amp.).

Hence the general procedure is to decrease (or increase) the unbalanced currents at those junctions where the unbalance is the greatest. Two columns should be established at every junction: (1) One column showing the change in the absolute potential assumed. (2) The other showing the new unbalanced currents (or voltages, or both).

There will be many more entries in the current column since the current varies also whenever the potential at a *neighboring* junction varies.

Figure 19 shows the steps necessary to reduce

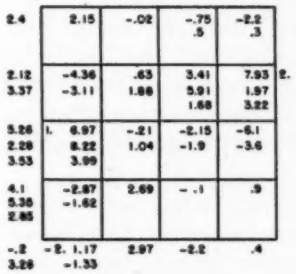


FIG. 19. Reduction of unbalanced currents to one-half of their value (from 8 percent to 4 percent).

in the example of Fig. 15 the unbalanced currents from 8 percent to 4 percent. *By changing the voltages only at five junctions, the accuracy of the a.c. Network Analyzer results has been increased 100 percent.* Figure 20 shows the steps necessary to reduce the unbalanced currents from 4 percent to 2 percent. Now 24 changes had to be assumed instead of five to increase the accuracy by another 100 percent. To reduce the unbalanced currents from 2 percent to 1 percent probably the same or greater number of changes would be necessary.

It is emphasized that a practiced person could probably have made the same reduction by less than 24 steps. Also, the network is symmetrical, a fact which was not considered in the reduction.

No hard and fast rules can be established for the reduction. While it is possible to assume at a junction a change of potential that reduces its own unbalanced current immediately to zero, it will not remain so as soon as the potential at a neighboring junction is assumed to vary. Often

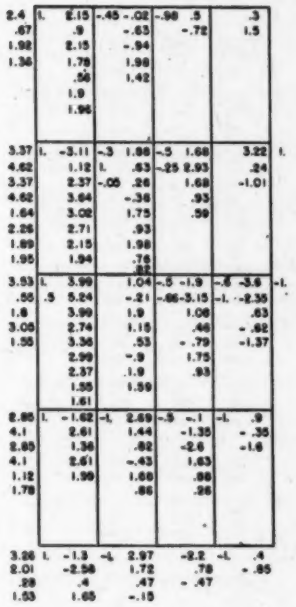


FIG. 20. Reduction of unbalanced currents to one-fourth of their value (from 4 percent to 2 percent).

it is advantageous to reduce an unbalanced current not only to zero but to a negative value.

Group Relaxation

There is no limit of course to the amount and type of labor-saving devices that can be introduced. One of them will now be considered.

Instead of assuming a change of voltage at one junction, a change of voltage at several junctions may be assumed simultaneously. This requires setting up "current patterns" for several junctions. An example for three border junctions is shown in Fig. 21.

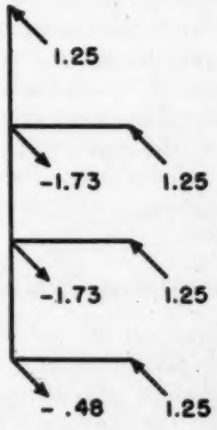


FIG. 21. Current pattern for three border junction.

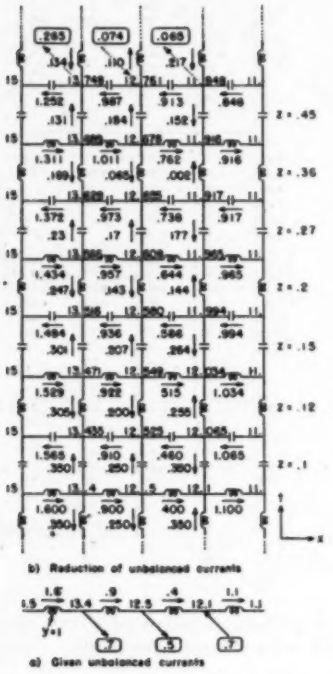


FIG. 22. The "Diffusion" method.

Networks with Complex Impedances

If both resistances and inductances occur in a network, then the unbalanced junction-currents are complex numbers i_1+ji_2, instead of real numbers i_1. Similarly the absolute potentials are complex numbers. Then a separate current pattern has to be established for a unit change in the *real* and the *imaginary* components of junction-voltage respectively. The latter must so change that both the real and imaginary parts of unbalanced currents approach zero.

Another and perhaps a better method is to replace the equivalent circuit by a more complex circuit in which only real numbers occur. Such a step is always possible since a set of differential equations containing complex coefficients may always be replaced by a set containing no j by simply doubling the number of equations and the number of dependent variables, that is, by introducing in-phase and out-of-phase components of variables.

The "Diffusion" Method

While the relaxation and the weighted average methods change potentials at one junction at a time, the method of unbalanced currents and voltages changes the potentials at all junctions simultaneously. The first two methods are found to be effective at the beginning of the reduction, especially in the elimination of "bumps" in otherwise smooth curves; afterwards the latter method appears to be faster. When the unbalanced currents are small compared with the circuit currents but still further reduction is desired, all three methods become very slow. The following method offers a further systematization of the third method.

A partial differential equation, not containing time, usually is considered to represent summation of currents at a junction. An unbalanced current i_u has no counterpart in the equation. Let it be assumed that i_u corresponds to an additional term of the form $A\partial\varphi/\partial t$ in the equation, adding thereby an additional independent variable t to the equation and an additional dimension to the network, such as shown in Fig. 4, for the wave equation.

From physical consideration of diffusion problems, it appears that if the initial conditions are well selected, the rate of diffusion decreases as time increases, and the currents representing $A\partial\varphi/\partial t$ must become in general smaller. At $t=\infty$ they all must become zero, allowing the network to represent the original equation without the time term.

By experience it has been found that at $t=0$, half of the unbalanced current should flow up, half down (Fig. 22 representing the one-dimensional wave equation). The coefficient A should be so selected that at first the potentials on the second layer should differ by only a small amount (say one percent or less) from the starting potentials. As time goes on, the value of A may be increased to speed up the rate of diffusion. The values of A should be so selected that the bigger currents between the layers should decrease uniformly along smooth curves instead of oscillate around them. (Some of the smaller currents will increase.) Of course A may be an arbitrary function not only of time but also of space, allowing thereby speedier diffusion at certain points.

In actual calculations, especially when the layers are two-dimensional, it is sufficient to have the drawing of one layer only and write the subsequent values of potentials in columns, as in the relaxation method.

PART III. CHARACTERISTIC-VALUE PROBLEMS

Oscillatory Circuits

In characteristic-value problems not only the mode of vibration (the junction-potential distri-

bution) is unknown, but also the frequency ω corresponding to that mode. If, however, a potential distribution is assumed, the corresponding ω may be found by the following property of the networks, a consequence of Rayleigh's Principle.

"When the network potentials correspond to a characteristic function, the algebraic sum of the positive and negative powers in the inductors and capacitors is zero."

That is, the power $\sum I^2X_L=\sum E^2Y_L=\sum EI_L$ in the inductors is equal to the power $\sum I^2X_C$ in the capacitors. Or the power in the variable units, that are functions of ω, is equal to the power in the constant units. This zero power in the network is utilized in finding characteristic values with the aid of a Network Analyzer. The admittances of the coils that are functions of ω are varied with ω until a generator, connected in shunt with any of the coils, draws no current from the line. At such values of admittances, the network is self-supporting and the stored electrical energy oscillates between the capacitors and the inductors.

Calculation of Characteristic Values

Hence, for an assumed potential distribution e, the corresponding ω is found by the following steps: 1. Calculate the power $\sum I^2X$ in all the coils that are not functions of ω (positive power exists in inductors and negative in capacitors). This power may also be found by calculating the power flowing in the coils that are functions of ω. If the latter are ground coils, then if e is the junction potential and i the current through the ground coils, power $=\sum ei$. 2. If the admittance of each of the remaining coils is the same and is equal to ω^2C, then the power in these coils is $e^2y=\sum e^2\omega^2C=\omega^2C\sum e^2$. Hence

$$y=\omega^2C=\frac{\sum I^2X}{\sum e^2}=\frac{\sum ei}{\sum e^2}.$$

By Rayleigh's principle it is well known that if the assumed mode of vibration (potentials) is only a rough approximation, the resultant characteristic value calculated is a better than rough approximation.

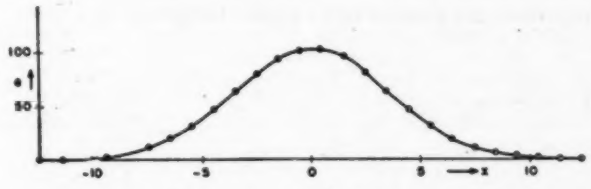


FIG. 23. A measured characteristic function of the linear harmonic oscillator.

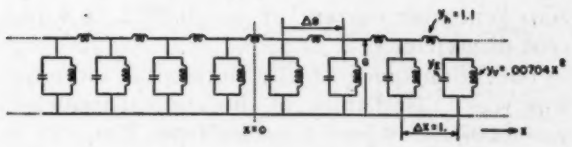


FIG. 24. The Schrödinger equation for the linear harmonic oscillator.

By knowing the correct ω for an assumed potential distribution, the correct unbalanced currents i_u and e_u may now be calculated. The latter in turn may be reduced by any of the four methods shown for the solution of boundary-value problems. However, in the present problems a new and more correct ω has to be calculated occasionally as the reduction proceeds.

The Method of Unbalanced Admittances

In the presence of coils that are functions of ω and whose admittances are usually equal, it is not necessary to introduce unbalanced currents. Instead the unbalance is redistributed by attributing different admittances to these variable coils. Hence when the correct ω has been calculated for a given potential distribution, it is assumed that *with each variable coil an unbalanced admittance y_u is associated*. The purpose of the reduction is then to reduce the values of these unbalanced admittances.

Expressed in another way, the purpose of the reduction is to make the admittances of all variable coils the same. The calculated ω only gives a goal to aim at at the beginning. As the reduction proceeds, the common value of the admittances continuously shifts.

The use of the "unbalanced current patterns" facilitates the calculation of the unbalanced admittances at the neighboring points, when the potential at a point is changed.

Linear Harmonic Oscillator

It will be assumed that an approximate value of a characteristic function of the linear harmonic oscillator is known, Fig. 23. (Actually it has been measured on a Network Analyzer,[9] with the aid of the network of Fig. 24.) The problem is to calculate the corresponding characteristic value. Theoretically that is known to be $y_E=0.0880$. The measurement on the Network Analyzer gave

zero generator current at $y_E=0.0921$, a 4 percent discrepancy.

The admittance of the horizontal inductors was $y_h=1.1$ and those of the vertical inductors $y_v=0.00704x^2$ where x varies from $\frac{1}{2}$ to $12\frac{1}{2}$ in steps of 1 as given in Table I for half the coils. The measured potentials from ground to junction are given in Table II.

TABLE I. Admittance of vertical coils.

0.176	8.62	29.7	63.5	110
1.583	14.24	39.6	77.6	
4.4	21.3	50.8	93.1	

TABLE II. Measured potentials e.

0.5	10.8	79	80.1	11	0.5
0.8	18.8	92.7	63.2	5.8	
1.5	30	100	46	3	
3	45.4	101.5	30.5	1.5	
5.8	62.4	94	19	0.8	

In the Schrödinger equation the admittance of the capacitors is proportional to ω. The energy in the horizontal coils is $\sum 1.1\Delta e^2=2851$, in the vertical coils it is $\sum e^2 y_v=2890$. Hence

$$y_E=\frac{\sum(e^2y_v+\Delta e^2\,1.1)}{\sum e^2}=\frac{5{,}741}{65{,}019}=0.0883.$$

This calculated energy level differs from the theoretical value of 0.0880 by only 3 parts in 10,000. It should also be noted that the true (and lowest) energy level 0.0880 is always *smaller* than the approximate levels 0.0883 or 0.0921, in accordance with Rayleigh's principle, or its quantum mechanical analogue

$$E<\frac{\sum\psi^*H\psi\Delta x}{\sum\psi^*\psi\Delta x}=\frac{\sum ei}{\sum e^2}.$$

Here ψ is the approximate wave function e. Also $H\Delta x\psi$ is the current i in the capacitor and $E\Delta x$ the admittance y_E.

BIBLIOGRAPHY

(1) H. W. Emmons, "Numerical methods of solving partial differential equations," Quart. App. Math. (Oct. 1944).
(2) A. Vazsonyi, "Numerical method in the theory of vibrating bodies," J. App. Phys. **15**, 598 (1944).
(3) G. Kron, "Equivalent circuits of the elastic field," J. App. Mech. **11**, 149–161 (1944).
(4) J. B. Scarborough, *Numerical Mathematical Analysis* (Oxford University Press, 1930).
(5) R. E. Doherty and E. Keller, *Mathematics of Modern Engineering I* (John Wiley & Sons, Inc., New York, 1936), p. 167.
(6) J. H. Bartlett, "The shaping of pole faces of the betatron," Phys. Rev. **63**, 185 (1943).
(7) G. Kron, "Equivalent circuits of compressible and incompressible flow fields," scheduled to appear in J. Aer. Sci.
(8) G. K. Carter, "Numerical and network analyzer solutions of the equivalent circuits for the elastic field," J. App. Mech. **11**, 162–167 (1944).
(9) G. K. Carter and G. Kron, "A.c. network analyzer study of the Schrödinger equation," Phys. Rev. **67**, 44 (1945).
(10) G. Kron, "Electric circuit models of the Schrödinger equation," Phys. Rev. **67**, 39 (1945).
(11) G. Kron, "Equivalent circuit of the field equations of Maxwell," Proc. I. R. E., pp. 289–299 (May, 1944).
(12) R. V. Southwell, *Relaxation Methods in Engineering Science* (Oxford University Press, 1940).
(13) A. F. Prebus, I. Zlotowsky and G. Kron, "The application of network analysis to some electron optical problems," scheduled to appear in Phys. Rev.

Erratum: Physical Aspects of Foaming in Steam Generation

[J. App. Phys. **15**, 792 (1944)]

HANS M. CASSEL
General Ceramics and Steatite Corporation, Keasbey, New Jersey

DR. Cassel's professional affiliation was incorrectly named on page 792 as General Ceramics and Acatite Corporation. It should have read "General Ceramics and Steatite Corporation."

Letter to the Editor

W. James Lyons
Southern Regional Research Laboratory, New Orleans, Louisiana

The birth of the profession of physics, as is disclosed, for instance, by the Report of the National Research Council Conference of Physicists,[1] is proving to be a very painful process, but one which is at least providing ample thought and discussion on the nature of physical research, and the place of the physicist in science and industry. While the various discussions published and reported during recent years in this Journal and the *Review of Scientific Instruments* have brought out numerous interpretations and opinions, they have failed to stress an historic tendency which, unique with physics, accounts for the present dispersed state of the profession, and has implications in the future well-being of physics as a profession. This is the tendency among physicists, and their forebears the natural philosophers, to lose further interest in a discovery, effect or process once its general, elementary principles have been mastered. The prevalence of such a tendency is reflected in much of the discussion at the Conference mentioned above. To expand on the reasons for this condition would be an unwarranted digression. Suffice it to say that the lack of interest in the utilitarian aspects comes much nearer being a professional characteristic in physics than it does in the research professions with which physics is most closely allied, *viz.*, chemistry and engineering.

Interpreted in certain ways (and unfortunately these interpretations have found wide favor among physicists), this behavior looms up as a rare gem, about which society ought cast all manner of protection. The physicists are pictured as men uniquely endowed with such insatiable curiosity, relentlessly driving them on, that they are unable to stop for a few years to further examine their findings, or explore their possible utility. With ideal scientific selflessness, physicists have, time after time, abandoned a newly found area of knowledge to leave its possible exploitation to others. The sustaining principle frequently brought into play at this point is that no body of newly discovered scientific knowledge is the exclusive preserve of any one group of men. It may be remarked that while this principle is entirely valid, it does not imply, as one would be inclined to conclude from developments in physical research, that the original explorers have no proper interests in the domain once its principal features have been disclosed.

Aside from the rather evident fact that no other professions accept this heroic, intraprofessional picture of the physicist, it should be noted that the "dilettante" attitude (as it has been called), which has prevailed in physics, does not "pay dividends." Interpreted realistically, the overemphasis of current fashions in physics, on the one hand, and the neglect of fields which have lost their novelty, on the other, are suicidal to a vital, well-supported profession. The reason this is so is that the funds for the support of science and research are distributed not by professional scientists, but by philanthropists, business men, legislators, and educational administrators. In general, these men are not prepared to judge the worth, or possible promise of future utility, of research at the raw frontiers of knowledge, of the type with which physics, in recent decades especially, has too frequently become preoccupied. Accordingly, the really substantial funds are allotted to those professions with which tangible, manifest benefits to society are associated. Repeatedly during the past one hundred years, however, by the time the benefits of some discovery or development in physics has become tangible and manifest, the physicists have departed, and in their stead is a new profession. Members of the new profession usually manage to conceive of their corner of physics as an independent science built from the ground up by the profession. By being neither affirmed nor denied, the debt to physics is effectively submerged. Physicists on their part, whether because of scientific snobbishness, or a frigid reception, or shyness in the presence of large scale commercial operations, avoid all but brief incursions of the field. The effect of the whole process, as has been indicated, is to divorce the profession of physics from a source of substantial and deserved support, no small part of which would be shared by pure, fundamental research.

The development of radio affords a representative example of the point being made here. Practically exclusively, the pioneer work was done by physicists: Kelvin, Maxwell, Hertz, Lodge, and Thomson, to mention the outstanding names. The term "physicist" as applied to these men is not being used in the loose sense of meaning anyone who is familiar with, and has applied some phase of, physical science. They were physicists according to the most rigorous definition, and by professional title, professors of experimental or natural philosophy, or of physics. Even after the inception of the art of wireless communication, during its infancy, the interest of physicists continued, as is attested by the publication in *The Physical Review* and *Annalen der Physik*, for instance, of early researches on the propagation and reception of electromagnetic radiation, electrical conduction in gases, etc. Unfortunately for physics, however, its association with radio progressively decreased as the utility and value of the art expanded. Today, though the war has created a need for him in communications research, the trained physicist does not know whether the fields of electrical and radio research provide permanently an appropriate professional outlet. It is, of course, trite to remark that current public opinion draws no connection between physics and radio. The lack of a professional association of physics with electrical and radio communication is in sharp contrast with the continued identification of chemistry with the industries which it has wholly or in part created.

The history of the development of radio seems to be well on the way to repetition with respect to the "science of electronics," as it has been called on a regular, coast-to-coast broadcast program. Nowhere amid all the high priced advertising, are the non-scientist managers and administrators told that electronic devices are contributions to better living through physics. It will be interesting to watch in the postwar years the professional trends accompanying further development of the cyclotron and the electron microscope. These instruments, invented and developed by *bona fide* physicists, thus far have remained principally associated with the physics profession. Shall we expect to read, however, say twenty years hence, that the cyclotron is the contribution and special care of the "cyclotronic engineers," banded together in an independent institute, and read perhaps, on the next page, that physicists are meeting to discuss means of initiating a professional attitude in physics? We may hope that the current discussions bring an end to past trends and lead physicists to a realization of their full professional heritage.

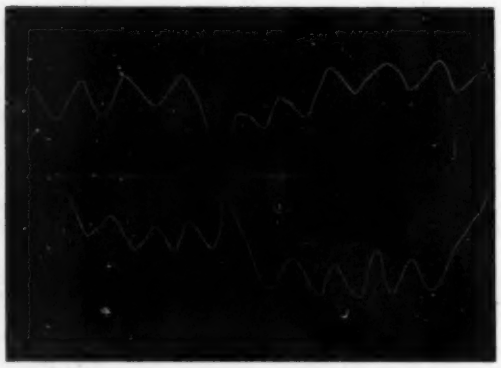
[1] Rev. Sci. Inst. 15, 283 (1944).

Recent Applications of Physics

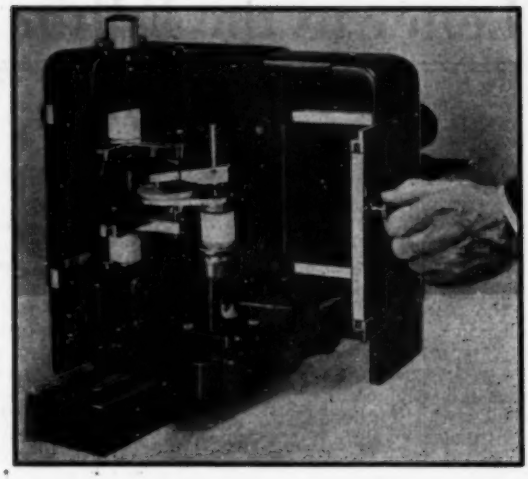
Prepared by CLARK GOODMAN, Associate Editor

Material to be included in this section should be submitted to Dr. Clark Goodman, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.

Fish Noises In order to study systematically the sounds made by fish, Dr. Christopher Coates of the New York Zoological Society Aquarium made oscillograms using an underwater microphone and a cathode-ray oscilloscope. The most immediate application of these results is to aid the crews of sub chasers to distinguish fish noises from the sounds of submarines. However, it is possible that such studies may be of value in aiding fishermen in the future. The accompanying photograph and oscillogram show a catfish and the soft purring sound which he makes when happy. Such sounds are involuntary nervous reactions, and there is no indication at present that these sounds are used for communication. Most fish noises are made by the grinding of teeth, while some fish blow air from swim bladders to make croaking noises.



Arc-Spark Stand The diverse applications of spectrographic analysis make a versatile type of source a necessity in many laboratories. The ARL-Dietert arc-spark stand shown in the photograph below attempts to fill this need. It combines into one unit



the features of the conventional arc stand, the jig-loaded stand used for metal-rod analysis and the Petrey stand used for metal-plate analysis. A friction drive accurately positions the electrodes. It provides a 1-inch continuous travel for both the upper and lower holders, a minimum spacing of $\frac{1}{8}$ inch between holders and a maximum of $2\frac{1}{8}$ inches. The motion of each electrode holder is limited by calibrated stops adjustable to 2 mils with a vernier.

The holders may also be locked in place and jigs used for spacing sample rods or rod-carbon electrodes. By replacing the upper electrode with a rotary table, pinned for exact positioning, sparking at three locations on the flat specimen can be carried out by merely moving an external lever. A 3-degree tilt of the table prevents shadowing of the discharge by the overhang of the sample.

Echolocation by Bats Aeronautical radar engineers may find an interesting and surprising parallel between their techniques and those used by bats. Recent studies, by R. Galambos and D. R. Griffin of Harvard University, have shown that bats are equipped with a supersonic system of echolocation that enables them to avoid obstacles while flying in the dark. The powerful larynx muscles and thin, short vocal cords enable the bat to emit intense bursts of sound at frequencies of 30 to 70 kc, the maximum intensity occurring about 50 kc. This sound is quite inaudible to the human ear, since the upper limit of human hearing is about 20 kc. However, bats appear to be unique among mammals in having extraordinary morphological peculiarities in the cochlea that enable them to hear supersonic sounds as high in frequency as 100 kc. The probing bursts last about 5 milliseconds and are emitted at varying rates of 5 to 10 per second when the bat is at rest, 20 to 30 per second in normal flight and up to 50 or 60 per second when approaching an obstacle. The difference in intensity or time of arrival of the reflected or scattered sound enables the bat to locate the obstacle and hence guide his flight in order to avoid a collision. Obstacles as small as a 16-guage wire can apparently be located in this way.

The use of random sound cues by blind men is well known. It is possible that a portable supersonic device whose echoes are converted into audible sounds heard only by the bearer would be of valuable assistance to the blind.